Stony Brook University

Toward Precision Medicine:
From Clinical Genomics to iPSC Disease Modeling

A Dissertation Presented

by

Yiyang Wu

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Genetics

Stony Brook University

May 2017

Stony Brook University
The Graduate School

Yiyang Wu

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Gholson J. Lyon - Dissertation Advisor
Assistant Professor of Genetics
Cold Spring Harbor Laboratory

Rolf Sternglanz - Chairperson of Defense
Distinguished Professor Emeritus of Biochemistry and Cell Biology
Stony Brook University

Alea A. Mills
Professor of Genetics
Cold Spring Harbor Laboratory

Emilia Entcheva
Professor of Biomedical Engineering
The George Washington University

Kristen Brennand - Outside Member
Associate Professor of Genetics and Genomics, Neuroscience, and Psychiatry
Icahn School of Medicine at Mount Sinai

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Toward Precision Medicine:
From Clinical Genomics to iPSC Disease Modeling

by

Yiyang Wu

Doctor of Philosophy

in

Genetics

Stony Brook University

2017

The ability to unveil the genetic landscape of human disease at an extraordinarily detailed resolution through high-throughput DNA sequencing promises a transformation of medical practice toward precision medicine. However, it is not a straight path from producing individual genomics data to pinpointing the disease-contributory variants and delivering the personalized therapy, due to the complexity of human genome architecture, genotype-phenotype correlation and disease pathogenesis. But first and foremost for precision medicine to succeed is to ensure the reliability and accuracy of the generation and interpretation of clinical genomics. However, due to a lack of an unbiased high-throughput targeted resequencing protocol, the validity and comparability of each DNA sequencing platform, protocol and variant detection algorithm has yet to be characterized. Meanwhile, a comprehensive investigation of the clinical findings must be equally executed in order to achieve a better design of sequencing projects and prioritization of disease-associated variants, especially for patients presenting a complex history. However, the hurdles to obtaining first-hand clinical data and a lack of standardized vocabulary use in medical documentation have prohibited this information from being utilized to its maximum potential. Furthermore, before any therapeutics can be developed, a robust disease model needs to be constructed to prove the causality of the pinpointed variants and understand the functional impact of it in disease affected cell types or tissues.

This dissertation research aims to address some of the above issues to help advance the implementation of precision medicine. In particular, the first study aims to assess the validity issue of clinical genomics data by providing a robust, high-throughput, targeted resequencing protocol. Study 2 aims to illustrate the opportunity and challenge of utilizing clinical genomics to identify the genetic basis of human diseases with two real-world examples, including our newly discovered genetic disorder TAF1 syndrome. Study 3 aims to investigate the functional impact of Naa10 S37P, a variant identified through clinical genomics that contributes to a rare disease Ogden syndrome, in patients' fibroblasts. And the last study aims to model the cardiac dysrhythmias of Ogden syndrome using patient specific induced pluripotent stem cell-derived cardiomyocytes to shed some light on the mechanism of Ogden syndrome pathogenesis.

Dedicated
to

The Pursuit of Knowledge
and The Health and Welfare of Mankind

Contents

0 Introduction **1**

1 Advancing the quality of next-generation sequencing **8**
- 1.1 Introduction . . . 8
- 1.2 Materials and Methods . . . 15
- 1.3 Results . . . 19
- 1.4 Discussion . . . 34
- 1.5 Concluding Remarks . . . 42

2 Accelerating discovery of the genetic basis of disease via clinical genomics **44**
- 2.1 Introduction . . . 44
- 2.2 Materials and Methods . . . 47
- 2.3 Results . . . 51
- 2.4 Discussion . . . 75
- 2.5 Concluding Remarks . . . 80

3 Ogden syndrome is linked to downstream N-terminal acetylation defects **81**
- 3.1 Introduction . . . 81
- 3.2 Materials and Methods . . . 86
- 3.3 Results . . . 94
- 3.4 Discussion . . . 114
- 3.5 Concluding Remarks . . . 130

4 Better understanding cardiac arrhythmia in Ogden syndrome via an iPSC model **132**
- 4.1 Introduction . . . 132
- 4.2 Materials and Methods . . . 141
- 4.3 Results . . . 152
- 4.4 Discussion . . . 184
- 4.5 Concluding Remarks . . . 194

5 Conclusions **196**

References **202**

List of Figures

1.1 General workflow of next-generation sequencing based on Illumina technology . 10
1.2 General workflow of data generation for targeted resequencing using Illumina MiSeq platform . . . 22
1.3 Agarose gel images of test PCR reaction during all three validation projects . . . 24
1.4 Agarose gel images of target PCR reaction during all three validation projects . . 27
1.5 Bioanalyzer analysis of the post library construction samples . . . 29
1.6 An example of high quality reads from both strands (BAM file) . . . 38
1.7 Two examples of reads with strand bias (BAM file) . . . 39
1.8 An example of reads with interesting patterns (BAM file) . . . 40
1.9 An example of reads with end problems (BAM file) . . . 41

2.1 Pedigree of the studied family in project 1 (the TAF1 story) . . . 53
2.2 Facial features, peculiar gluteal crease and MRI images of probands discussed in project 1 . . . 54
2.3 Additional phenotypic images of probands discussed in project 1 . . . 55
2.4 Sanger sequencing analysis of 10 family members in project 1 for variants found in *ZNF41*, *ASB12*, *PION* and *TAF1* . . . 62
2.5 Pedigree of the studied family in project 2 . . . 66
2.6 Sanger sequencing results of the variant found in *POU4F1* . . . 78

3.1 Ogden syndrome pedigrees from two unrelated families with eight probands . . 83
3.2 Sanger sequencing analysis of NAA10 S37P allele in primary fibroblasts . . . 101
3.3 Reduced NatA complex formation of NAA10 S37P . . . 103
3.4 Altered growth rate, proliferation and metabolic activity of NAA10 S37P fibroblasts . . . 105
3.5 Mitochondrial structure evaluation of S37P primary fibroblasts . . . 108
3.6 Real-time mitochondrial respiration assessment of NAA10 S37P fibroblasts . . . 111
3.7 Real-time glycolytic activity assessment of NAA10 S37P fibroblasts . . . 113
3.8 Human pathogenic mutations identified in NAA10 . . . 115
3.9 Tissue-specific expression pattern of NAA10 and NAA15 in histologically normal human samples . . . 123

4.1 ECG analysis of an Ogden syndrome patient . . . 153

4.2 Generation of RNA-induced pluripotent stem cells from primary fibroblasts . . . 156
4.3 Profile of NatA substrates in mouse proteome during reprogramming 159
4.4 Generation of Sendai virus-induced pluripotent stem cells from primary fibroblasts 161
4.5 Characterization of human iPSC line WT-1 #1, #2, #5 and S37P-D, I, H 163
4.6 Generation of iPSC-derived cardiomyocytes using “GiWi” protocol 167
4.7 Immunostaining analysis of iPS-CMs at 5 months old 168
4.8 Spontaneous arrhythmic activities of 3 months old S37P iPS-CMs 171
4.9 Optogenetically modified iPSC-CMs express ChR2 173
4.10 Membrane potential recording of 5- and 6-month old iPS-CMs by optical stimulation 176
4.11 Spontaneous arrhythmic activities of 3-month old S37P-H iPS-CMs during calcium transient recording 179
4.12 Intracellular calcium transient evaluation of 3-month old iPS-CMs by electrical stimulation 181
4.13 Intracellular calcium transient evaluation of 6-month old iPS-CMs by optical stimulation 183
4.14 Varying membrane potential durations of 3-month old iPS-CMs by electrical stimulation 188

List of Tables

1.1 Technical comparison of popular next generation sequencing platforms 12
1.2 MiSeq validation data summary 31
1.3 MiSeq validation rate for SNVs/Indels specific to each pipeline/platform 32

2.1 Clinical features of probands in the TAF1 story 56
2.2 Clinical presentation of proband K10031-10232 68
2.3 Clinical presentation of proband K10031-10133's family 71

3.1 Clinical features of Ogden syndrome 97
3.2 Overview of protein N termini less acetylated in S37P cells and siNatA HeLa cells 125

4.1 Cardiac features of Ogden syndrome 135
4.2 Predicted targets of NatA linked to channelopathy diseases 138

List of Abbreviations

APD	action potential duration
ASD	atrial septal defect
bp	base pair
CHD	congenital heart disease
CM	cardiomyocyte
CNV	copy number variation
CPVT	catecholaminergic polymorphic ventricular tachycardia
CTD	calcium transit duration
EB	embryoid body
ECAR	extracellular acidification rate
ECG	electrocardiogram
EP	electrophysiology
ERC	electrical restitution curve
FD	familial dysautonomia
HDFn	human dermal fibroblasts, neonatal
(h)ESC	(human) embryonic stem cell
HH	hereditary hemochromatosis
(h)iPSC	(human) induced pluripotent stem cell
HPO	the human phenotype ontology
(h)TERT	(human) telomerase reverse transcriptase

IF	immunofluorescence
INDEL	insertion or deletion
IP	immunoprecipitation
iPS-CM	iPSC-derived cardiomyocyte
LQTS	long QT syndrome
MRI	magnetic resonance imaging
NAT	N-terminal acetyltransferase
NGS	next-generation sequencing
NTA	N terminal acetylation
OCR	oxygen consumption rate
OXPHOS	oxidative phosphorylation
PAC	premature atrial contraction
PCR	polymerase chain reaction
PDA	patent ductus arteriosus
PFO	patent foramen ovale
PTM	post-translational modification
PVC	premature ventricular contraction
PWS	Prader-Willi syndrome
SeV	Sendai virus
SNV	single nucleotide variation
SVT	supra-ventricular tachycardia
TDP	Torsade de Points
VSD	ventricular septal defect
VT	ventricular tachycardia
WES	whole exome sequencing
WGS	whole genome sequencing
WT	wild-type
XCI	X-chromosome inactivation

Acknowledgments

I would like to thank my dissertation advisor Dr. Gholson Lyon for giving me the opportunity to study and work at Cold Spring Harbor Laboratory and the freedom to explore in science. I would like to thank my dissertation co-advisor Dr. Emilia Entcheva for her mentorship and support on the cardiac research I conducted in her lab at both Stony Brook and Washington DC. I would like to thank my dissertation committee, Dr. Rolf Sternglanz, Dr. Alea Mills, Dr. Kristen Brennand, and former members Dr. Chris Vakoc and Dr. Geoffrey Girnun for all of their guidance, patience and encouragement over the past few years. Their expertise and wisdom have helped me grow faster as a young scientist.

Without the help from my research collaborators Aleks Klimas, Dr. Sunita D'Souza, Dr. Geoffrey Girnun, Dr. Christoph Schaniel, Dr. Jason A. O'Rawe, Han Fang, this extremely challenging dissertation work would not be possible. So I thank them deeply.

I would like to thank each of my co-authors on the studies mentioned in this dissertation, their dedication and expertise resulted in much stronger scientific papers.

I would like to thank the patients and their family members that I have studied for their understandings and sacrifice.

I would like to thank all of my teachers and science heroes in the past, for their foundational works and being my inspirations to become a scientist.

I would like to thank my entire medical care team, my family and friends, for their care and love, to keep me alive and function, and to make my journey in this world worthwhile.

Lastly and most importantly, I would like to extend an enormous thanks to my dearest mom and dad for being my rock and role models. I would not be here and could hardly have achieved anything without them.

Publications

Chapter 1 Related

Fang, H., Bergmann, E., Arora, K., Vacic, V., Zody, M., Iossifov, I., O'Rawe, J., **Wu, Y.**, Jiménez-Barrón, L. T., Rosenbaum, J., Ronemus, M., Lee, Y., Wang, Z., Dikoglu, E., Jobanputra, V., Lyon, G. J., Wigler, M., Schatz, M., & Narzisi, G. (2016). *Indel variant analysis of short-read sequencing data with Scalpel*. Nature Protocols, 11(12):2529-2548.

Fang, H., **Wu, Y.**, Narzisi, G., O'Rawe, J., Jiménez-Barrón, L. T., Rosenbaum, J., Ronemus, M., Iossifov, I., Schatz, M., & Lyon, G. J. (2014). *Reducing INDEL calling errors in whole genome and exome sequencing data*. Genome Medicine, 6(10):89.

Narzisi, G., O'Rawe, J. A., Iossifov, I., Fang, H., Lee, Y.-h., Wang, Z., **Wu, Y.**, Lyon, G. J., Wigler, M., & Schatz, M. C. (2014). *Accurate de novo and transmitted indel detection in exome-capture data using microassembly*. Nature Methods, 11(10):1033-1036.

O'Rawe, J, Jiang, T., Sun, G., **Wu, Y.**, Wang, W., Hu, J., Bodily, P., Tian, L., Hakonarson, H., & Johnson, W. E. (2013). *Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing*. Genome Medicine, 5(3):28.

Chapter 2 Related

Fang, H.*, **Wu, Y.***, Yang, H., Yoon, M., Jiménez-Barrón, L. T., Mittelman, D., Robison, R., Wang, K., & Lyon, G. J. (2017). *Whole genome sequencing of one pedigree illustrates the challenges with genomic medicine*. BMC Medical Genomics. 10(1):10.

Jiménez-Barrón, L. T., O'Rawe, J. A., **Wu, Y.**, Yoon, M., Fang, H., Iossifov, I., & Lyon, G. J. (2015). *Genome-wide variant analysis of simplex autism families with an integrative clinical-bioinformatics pipeline*. Cold Spring Harbor Molecular Case Studies, 1(1):a000422.

O'Rawe, J. A., **Wu, Y.**, Dorfel, M. J., Rope, A. F., Au, P. Y., Parboosingh, J. S., Moon, S., Kousi, M., Kosma, K., Smith, C. S., Tzetis, M., Schuette, J. L., Hufnagel, R. B., Prada, C. E., Martinez, F., Orellana, C., Crain, J., Caro-Llopis, A., Oltra, S., Monfort, S., Jimenez-Barron, L. T., Swensen, J., Ellingwood, S., Smith, R., Fang, H., Ospina, S., Stegmann, S., Den Hollander,

N., Mittelman, D., Highnam, G., Robison, R., Yang, E., Faivre, L., Roubertie, A., Riviere, J. B., Monaghan, K. G., Wang, K., Davis, E. E., Katsanis, N., Kalscheuer, V. M., Wang, E. H., Metcalfe, K., Kleefstra, T., Innes, A. M., Kitsiou-Tzeli, S., Rosello, M., Keegan, C. E., & Lyon, G. J. (2015). *TAF1 Variants Are Associated with Dysmorphic Features, Intellectual Disability, and Neurological Manifestations*. American Journal of Human Genetics, 97(6):922-932.

*Equal Contribution

Chapter 3 Related

Myklebust, L., Van Damme P, Støve, S., Dörfel, M., Abboud, A., Kalvik, T., Grauffel, C., Jonckheere, V., **Wu, Y.**, Swensen, J., Kaasa, H., Liszczak, G., Marmorstein, R., Reuter, N., Lyon, G. J., Gevaert, K., & Arnesen, T. (2015). *Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects*. Human Molecular Genetics, 24(7):1956-1976.

Chapter 4 Related

Klimas, A., **Wu, Y.**, Ambrosi, C., Yu, J., Williams, J., Bien, H., Lyon, G. J., & Entcheva, E. (2016). *Disease Modeling in Human Induced Pluripotent Stem Cell Derived Cardiomyocytes Using High-Throughput All-Optical Dynamic Cardiac Electrophysiology*. Frontiers in Optics 2016, OSA Technical Digest (online) (Optical Society of America, 2016), paper FF3A.3.

Wu, Y., Klimas, A., D'Souza, S., Schaniel, C., Girnun, G., Entcheva, E., & Lyon, G. J. (2016). *Protein N-terminal Acetylation is Associated with Cardiac Rhythm Regulation*. Heart Rhythm, 12(5S):97-98.

Portions of this thesis (Chapter 3 & 4) contain manuscripts in preparation (not listed). Permission to publish, exhibit, or broadcast materials from these chapters must be requested in writing. This publication list was last updated on 5/12/2017.

Chapter 0

Introduction

Every single patient is different, and doctors have recognized it for a long time in their practice. The earliest documented recognition of individual variability might be from Dr. William Osler, who wrote in his influential textbook *The Principles and Practice of Medicine*, "If it were not for the great variability among individuals, medicine might as well be a science, not an art." (Osler, 1892). Variations among people can be easily identified through things such as eye colors, height, weight, blood pressure, and body temperature, and can also be reflected systematically by how we uniquely respond to disease and treatment. Fundamentally, these individual variations are caused by the underlying differences coded in our genome, which can

also be influenced by our environmental factors. The first person who recognized the role that genetics plays in determining human's variable traits is probably Archibald Garrod, who is now considered the founder of medical genetics. In his work (Garrod, 1902), Dr. Garrod illustrated that a rare disease alkaptonuria is inherited according to Mendelian rules (Weldon, 1902). Based on this work, he shared a very important insight in his well-known book *Inborn Errors of Metabolism* (Garrod, 1909): "these [alkaptonuria, albinism, and cystinuria] are merely extreme examples of variations of chemical behavior which are probably everywhere present in minor degrees and that just as no two individuals of a species are absolutely identical in bodily structure neither are their chemical processes carried out on exactly the same lines."

Realizing the existence of variations among individuals, Dr. Garrod later wrote in his book *The Inborn Factors in Disease* (Garrod, 1931): "The task of the practitioner is far more than to apply the knowledge supplied to him from the laboratories; he ... calls upon his experience to guide him as to how he may best help the particular patient [manage his disease] with the least possible damage". This seems to be the first clear statements of the goals of precision medicine (Perlman & Govindaraju, 2016), a model of health care that takes individual variability into account. However, due to a lack of reliable measurements for individual variation, modern medicine has been taking a "one-size-fits-all" approach where most of the treatments are designed for the "average patient". This has proven to be a very successful strategy in some patients, but at the same time causes many troubles for others.

The ability to identify human genetic variation and to investigate the precise genetic contributors of human disease has historically been very limited. In the case of alkaptonuria, it took 56 years before the defective enzyme was narrowed down to the homogentisate 1,2 dioxygenase (HGD) (La Du et al., 1958). And almost a century after Dr. Garrod published his

theory, the first HGD mutations in humans were demonstrated (Fernández-Cañón et al., 1996). However, this landscape of genetic research has been changed over the past decade since the advent of many emerging technologies such as high throughput genome-scale DNA sequencing technologies (next-generation sequencing, NGS), which enables researchers to efficiently study the genetic basis of diseases at nucleotide resolution. Such advancement also starts to shift the model of clinical practice from "one disease-one treatment" to genomic medicine, where treatment or health management plans are designed based on patients' unique genetic predisposition, as Dr. Garrod's proposed in 1931. The US Precision Medicine Initiative (PMI) was launched under this background in January 2015 by President Obama (Office of the Press Secretary, 2015) aiming to accelerate this process.

It is estimated that 6,000-7,000 different types of monogenic diseases (or rare diseases) have been identified to date, with more than half of them being successfully linked to a rare variant (minor allele frequency (MAF) < 1%) (Boycott et al., 2013). Whereas in the pathogenesis of complex diseases, multiple genetic variations, common (MAF > 5%), low frequency (MAF 1-5%) and/or rare, are believed to interact with environmental factors together (Arnar & Palsson, 2017; Cirulli & Goldstein, 2010; Lowe & Reddy, 2015; Mitchell, 2012). But regardless of the differences in genetic complexity, the application of precision medicine has been proven effective for diseases under both categories. Some well-known examples that have demonstrated the promise of precision medicine include a recent FDA approved drug therapy ivacaftor/lumacaftor for treating cystic fibrosis patients carrying homozygous CFTR F508del mutation (Lopes-Pacheco, 2016; Martiniano et al., 2016), and the development of personalized therapy for lung adenocarcinoma patients: gefitinib for EGFR mutation carriers and crizotinib for ALK mutation carriers (Blumenthal et al., 2016; Lindeman et al., 2013).

Another promising field of the application of precision medicine is pharmacogenomics (PGx), which studies the genetic basis of variability in drug response (Roses, 2000). The ultimate goal of pharmacogenomics is to realize "customized" prescriptions: to offer the right treatment to the right person at the right dose, to maximize the therapeutic drug effect, while minimizing adverse reactions and avoiding toxicity. There are several well-known examples of personalized medicine applications based on using PGx (Dickmann & Ware, 2016; Zhang & Nebert, 2017), one early but classic example is to screen HIV patients for HLA-B*5701 allele before abacavir treatment to avoid life-threatening hypersensitivity reactions (Hughes et al., 2009; Mallal et al., 2008; Martin & Kroetz, 2013), which affect about 5-8% of all patients (Hetherington et al., 2001). This genetic screening has been proven to be very cost-effective (Nieves Calatrava et al., 2010), hence, it has been routinely prescribed in some country such as UK.

All the above examples elucidate the key to the concept of precision medicine, which is to tailor treatments based on an individual's genetic makeup. However, none of these would be made possible without a detailed understanding of the molecular etiology of diseases, which is a crucial component to achieve precision medicine (Collins & Varmus, 2015). Unlike relatively easy access to patients' blood samples for genome analysis, it is extremely difficult to access the major disease-affected tissues (e.g. brain and heart cells) that are also expandable for mechanistic study and/or drug screening. Animal models such as mice, rats and non-human primates are all invaluable tools for studying human diseases due to the high evolutionary conservativity among mammalian genomes; however, the developmental and physiological differences between these species and humans are also substantial (Chinwalla et al., 2002; Hamlin & Altschuld, 2011). The advent of human induced pluripotent stem cell (hiPSC) technology offers a great opportunity to solve some of the aforementioned problems.

It has been 10 years since hiPSC technology was first discovered (K Takahashi, 2007; J. Yu et al., 2007), numerous technical advances have been made in this field, including refined protocols for hiPSC reprogramming, hiPSC differentiation into various cell types and organoids, and the integration of genome editing strategies in hiPSCs (Hotta & Yamanaka, 2015; Karagiannis et al., 2016; Ohnuki & Takahashi, 2015; Shi et al., 2016). iPSC technology enables the reprogramming of a somatic cell into a pluripotent stem cell without altering the genetic characteristics of an individual patient, which is achieved by applying four transcription factors - OCT4, SOX2, KLF4 and MYC. The biggest advantages of using hiPSCs for disease modeling lie in their human (patient) origin and potential to give rise to almost all cell types desired. Most of the successful hiPSC disease modeling studies to date focus on early onset monogenic disorders that present a clear disease phenotype, whereas in general it is much harder to model a complex disease, especially late onset, caused by predominant epigenetic or environmental factors.

So far, a broad range of diseases have been studied using disease-relevant hiPSCs derivatives (Young et al., 2012), which led to valuable new mechanistic insights. Take cardiac diseases as examples (which will be discussed more in chapter 4 of this dissertation): iPSC-derived cardiomyocytes (iPS-CMs) (Yazawa et al., 2011) have been used to model Timothy syndrome (long QT syndrome 8), a multisystem disorder caused by gain-of-function mutations in the L-type calcium channel *CACNA1C* gene (Splawski et al., 2004). The phenotype associated with Timothy syndrome iPS-CMs includes irregular and slow contraction, prolonged action potentials and delayed afterdepolarization, and larger and prolonged Ca^{2+} transients; iPS-CMs and iPSC-derived myocardium ("heart-on-chip") have successfully recapitulated the cardiac phenotype of X-linked monogenic mitochondrial cardiomyopathy Barth syndrome (G. Wang et

al., 2014). An impaired sarcomere assembly, abnormal mitochondrial function, and an increased mitochondrial reactive oxygen species production were seen from the patients' iPS-CMs, and the cardiac micro tissues generated from the patients' iPS-CMs also showed contractile dysfunction. Furthermore, 3D-engineered human cardiac organoids were recently developed and used to model acute cryoinjury, and the intrinsic regenerative potential of this immature heart tissue (which resembles fetal heart tissue according to the authors) was revealed (Voges et al., 2017).

With no doubt, the realization of precision medicine in the genomics era will bring infinite promises and potential years down the road (Francis S. Collins, 2014); however, there are still many major hurdles that must be overcome for a real-world implementation. Some of the current challenges include: the analytical and clinical validity of genomic tests has yet to be fully evaluated, majority of disease-associated variants identified do not have proven causality, molecular etiology of many diseases are still unclear due to a lack of robust disease models especially for complex diseases, and many regulatory barriers and ethical challenges prohibit the health-related information to be shared and exchanged (Ashley, 2016; Dong et al., 2015; Edwin, 2016; Kohane, 2015; Korngiebel et al., 2017; Pettitt et al., 2016). In this dissertation work, I will discuss four separate studies that aim to assess and improve some of the above challenges to help move the field forward.

The first study aims to develop a robust and high throughput validation protocol for assessing the reliability and accuracy of different next-generation sequencing platforms and variant detection tools, which has become the key for DNA sequencing technologies to be reliably used for routine medical care (Altman et al., 2016; Ashley, 2016; Kohane, 2015). Three projects will be introduced as examples of the application of such a protocol and to evaluate the

robustness and efficiency of it. Discordance of NGS variant calling will also be assessed and discussed.

The second study aims to explore the usefulness and practical limitations of using clinical-grade individual whole genomes to facilitate the discovery of the genetic basis of human diseases. Two real-world examples will be discussed, with one focused on a newly discovered disorder involving pathogenic TAF1 variants, and the other based on a complex pedigree with various medical conditions including Prader-Willi Syndrome (PWS), Hereditary Hemochromatosis (HH), and dysautonomia-like symptoms. Some regulatory barriers that influence the research outcome will be also discussed.

The third study aims to contribute to our understanding of the cellular and molecular consequences of a pathogenic variant, NAA10 S37P, found in the rare genetic disorder Ogden syndrome. An in-depth analysis of the phenotypic data of Ogden syndrome was performed. Parameters of cell growth, proliferation, protein interaction, metabolism and bioenergetics were evaluated using a NAA10 S37P cellular model. Genotype-phenotype correlations in NAA10-related disorders and a detailed molecular path to patients' phenotype will be discussed.

The final study aims to enhance our knowledge about cardiac arrhythmia pathogenesis by modeling the cardiac phenotype of Ogden syndrome patients in-a-dish using patient-specific induced pluripotent stem cell-derived cardiomyocytes. The electrophysiology properties of Ogden patient-derived cardiomyocytes were characterized in detail using an all-optical cardiac electrophysiology platform OptoDyCE. And insights of using patient-derived cardiomyocytes as a model to study heart disease will be discussed.

Chapter 1

Advancing the quality of next-generation sequencing

1.1 Introduction

Since the completion of the Human Genome Project in 2003 (Collins et al., 2003; International Human Genome Sequencing Consortium, 2004), tremendous advancements have been achieved in high-throughput DNA sequencing technologies, leading to a great drop of the cost and time to sequence a genome. These advancements have enhanced the discovery of breakthroughs in basic sciences, and also transformed modern medicine toward precision medicine (Goodwin et al., 2016; Huang et al., 2016; Koboldt et al., 2013; Van Dijk et al., 2014). However, the rapidly increasing need for the clinical applications of high-throughput DNA

sequencing, the lack of standards and comparison data for the performance of different sequencing platforms and multiple data-analysis pipelines has led to concerns in the field (Altman et al., 2016). It has become evident that the assurance of the production of reliable individual genomes and the accurate identification of genomic variants are necessary to leverage the power of high-throughput DNA sequencing to revolutionize patient care (Ashley, 2016; Basho et al., 2015; Gargis et al., 2016; Im et al., 2016; Kohane, 2015; Luthra et al., 2015; Roy et al., 2016; Solomon, 2014).

The production of a personal genome by NGS technologies contains three stages, including data generation, processing, and analysis, each of which consists of multiple steps (Fig 1.1). Due to the complexity of the process and the data structure, successful generation of high quality genomics data often requires multidisciplinary collaborations and relies on strict quality control of each step at all stages. However, assessing the quality of the data is not as straightforward as evaluating the quality of a DNA sample by spectrophotometers; it requires a deep understanding of the sources that can contribute to the quality uncertainties.

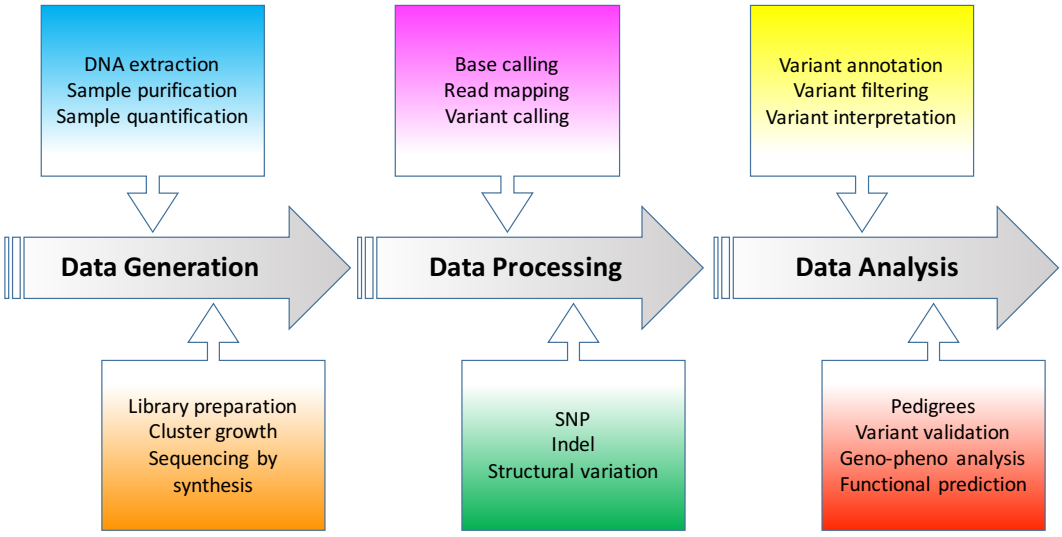


Figure 1.1 General workflow of next-generation sequencing based on Illumina technology. SNP, single-nucleotide polymorphism; Indel, insertion and deletion; Geno-pheno, genotype-phenotype.

Diversity of next-generation sequencing platforms

Following the human genome project, next-generation sequencing technology has been evolving rapidly. Multiple sequencing platforms have been launched in the past decade, among which Ion Torrent (Life Technologies), Pacific Biosciences (PacBio), Complete Genomics, Illumina Hiseq and MiSeq sequencers were some of the most popular NGS platforms used in the field during the time when I just started this project (in 2012). As seen, each platform has its unique merits as well as disadvantages in application due to different sample preparation protocol, sequencing chemistry, and signal detection technology being used (Table 1.1) (Lee et al., 2013; Liu et al., 2012; Mardis, 2013, 2017; Metzker, 2010; Quail et al., 2012). The existence of a variety of sequencing platforms gives researchers opportunities to select the "best" performance platform possible according to their own applications to achieve the most cost-effective run. However in the meantime, it also poses great challenges to compare data between groups if they do not share the use of the same sequencing platforms to begin with, since guidelines are missing for how to assess the differences contributed by various sequencing platforms, which can only be made possible by compared the true variants generated by each platform when applying the same samples.

Table 1.1 Technical comparison of popular next generation sequencing platforms

Platform	Illumina MiSeq	Illumina HiSeq 2000	Ion Torrent PGM	PacBio RS	Complete Genomics
Chemistry	sequencing by synthesis, reversible terminator	sequencing by synthesis, reversible terminator	sequencing by ligation, semiconductor technology, proton ion sensitive transistor	real-time, single molecule DNA sequencing	sequencing by ligation
Amplification/Library	bridge PCR, clusters, pair-end sequencing	bridge PCR, clusters, pair-end sequencing	emulsion PCR, pair-end sequencing	none	rolling circle amplification-DNA nanoball
Signal detection	fluorescent-dNTP	fluorescent-dNTP	ISFET ion sensor, CMOS semiconductor chip	fluorescent-dNTP	four fluorescently labeled 10-mer base probes
Max read length	2×150 bp	2×150 bp	200 bp	10 kb (average 1500 bp)	2×35 bp
Typical DNA requirements	50-1000 ng	50-1000 ng	~1 μg	100-1000 ng	≥ 5 μg
Sequence yield per run	120 Mb-1.5 Gb	600 Gb	20-50 Mb on 314 chip, 100-200 Mb on 316 chip, 1Gb on 318 chip	100 Mb-13 Gb	300-350 Gb (40× coverage), 20-60 Gb (variations only)
Run time	27 hours	2-11 days	2-8 hours	2 hours	12 days
Reported accuracy	> 99%	> 99%	> 99%	84~85%	99.999% for 40× depth
Observed raw error rate*	0.8%	0.26 %	1.71 %	12.86 %	N/A
Pros	high throughput, high coverage, short run time	high throughput	short run time, direct measurement of nucleobase incorporation events	short run time, very long read length	the sequencing reactions are non-progressive, low error rate
Cons	short read length, substitution errors between A/C and G/T	short read length, long run time	higher error rate in homopolymers, short read length	highest error rates compared with other NGS, no paired end reads	short read length, long run time, multiple rounds of PCR needed, hence high PCR bias and possibly amplify contaminants in the template construction phase

ISFET, ion-sensitive field-effect transistor; CMOS, complementary metal-oxide-semiconductor; NGS, next generation sequencing; bp, base pair; Mb, mega base pairs; Gb, giga base pairs; N/A, not available. *, (Quail et al., 2012).

Whole genome sequencing vs. whole exome sequencing

The most routinely used NGS technologies nowadays for clinical diagnostics is gene panel analysis due to its relatively low costs, short turnaround time and low rate of unspecific or incidental findings. It has been extremely useful as a rapid first-tier screening test to confirm or rule out a specific diagnosis of a known disease. However, when a negative result from gene panel analysis is given or an unknown genetic disorder is in doubt, it is time to call for a more comprehensive analysis by using either whole genome sequencing (WGS) or whole exome sequencing (WES). Both of them are effective tools to detect the disease-associated variants in rare and complex diseases in humans (Bamshad et al., 2011; Bolze et al., 2013; Alexandre Bolze et al., 2010; Byun et al., 2010; Genome of the Netherlands Consortium et al., 2014; Gilissen et al., 2014; Ng et al., 2009; Saunders et al., 2012; Tennessen et al., 2012), but each of them also has its own hurdles to overcome, such as problems with insufficient capture and missing meaningful variants outside the target regions for WES (Eisenberger et al., 2013; Meienberg et al., 2015), and a high cost and long turnaround time including data analysis for WGS. Hence, many outstanding questions such as when to use WGS or WES (e.g. what unique sequencing biases are related with each platform), and how to use them (e.g. how deep coverage is enough to generate a reliable genome) was still in debate, especially in the year of 2014, when project 3 discussed in this chapter was just initiated.

Multiple data-analysis pipelines

Compared to the variety of upstream data generation offered by different sequencing platforms, we face an even larger pool of bioinformatics tools to use for the downstream analysis from read alignment to variant calling. Representative read aligners include Bowtie2 (Langmead

& Salzberg, 2012), BWA mem (Li, 2013), MOSAIK (Lee et al., 2014), and Novoalign, and commonly used variant callers for detecting single nucleotide variants (SNVs) and short insertion-deletion polymorphisms (Indels) include GATK (McKenna et al., 2010), SAMtools (Li et al., 2009), FreeBayes (Garrison & Marth, 2012), SNVer (Wei et al., 2011), GNUMAP (Clement et al., 2009), SOAPsnp (Li et al., 2009), GATK-UnifiedGenotyper (DePristo et al., 2011), and Scalpel (Narzisi et al., 2014). Regardless of the fact that each tool was claimed to provide a better performance than others in one way or another, a systematic comparison of these analysis pipelines was rare in 2013.

The combination of above sources can give rise to quality uncertainties in the NGS data output, which can greatly compromise to a great extent the validity of using NGS-based clinical genomics as evidence for clinical decision making, which oftentimes leads to life-changing diagnoses and/or treatments for patients and their families. However, in order to address these above issues, first and foremost, we need to develop a robust, high-throughput, cost-effective validation protocol to facilitate the identification of true positive (TP) and true negative (TN) variants. There had been reports using genotyping arrays as a high-throughput validation method for evaluating variants generated by NGS (Lam et al., 2012), however, since microarrays tend to focus more on relatively common variants due to the limitation in the number of SNPs they can contain, it can be problematic to rely on them to validate unreported variants or variants with lower minor allele frequency (MAF) which are often seen in personal genomes from the patients.

1.2 Materials and Methods

Ethics approval

The sample collection and genomic analysis of DNA were approved by the institutional review boards at the University of Utah and Cold Spring Harbor Laboratory. Written informed consent was obtained from all study participants. SFARI (http://sfari.org/) maintains the consent of all individuals in the Simons Simplex Collection (SSC). Research was carried out in compliance with the Helsinki Declaration.

Sample collection

This process was carried out by other people. The samples used in our study came from families of human research participants ascertained in clinics at the University of Utah (for Project 1), and family quartets from the Simons Simplex Collection (SSC) (for Project 2 and 3) (Iossifov et al., 2012). DNA extracted from blood samples was quality-checked on agarose gels and quantified using NanoDrop 2000 (Thermo Fisher Scientific).

Whole genome/exome sequencing and microarrays

My role here was to prepare samples sent out for sequencing, and I was also involved with library construction (including all steps before bioanalyzer analysis). For Project 1, one in-house DNA sample K8101-49685s (10 μg) were sent to Complete Genomics (CG) (Mountain View) for WGS and analysis using CG 2.0 pipeline (Carnevali et al., 2012). DNA samples were also genotyped on Human610-Quad version 1 SNP arrays (Illumina) with approximately 610,000 markers at the Center for Applied Genomics (Children's Hospital of Philadelphia). The

exome was captured (I was not involved) using the Agilent 44 Mb SureSelect version 2 capture kit (Agilent Technologies) and then sequenced with Illumina HiSeq2000 paired-end reads (the platform used for all projects), averaging 100 bp in length after trimming. For project 3, the exome capture was performed (I was not involved) using 36 Mb NimbleGen SeqCap EZ Exome v2.0 kit (Roche). We also sent an in-house DNA sample K8101-49685s for WGS. (The individual of K8101-49685s has a severe case of Tourette syndrome and obsessive-compulsive disorder.) Both sets of libraries were sequenced on Illumina HiSeq2000 at the sequencing center of Cold Spring Harbor Laboratory.

Targeted resequencing using MiSeq

Variants were selected by Jason O'Rawe (project 1, 2), Han Fang (project 2, 3), and Giuseppe Narzisi (project 2). Primers were designed by Jason O'Rawe and myself. I was also involved in the study design.

For Project 1 discussed in the result section, 1,140 SNVs and 960 Indels found in sample k8101-49685s were randomly selected for validation. Among 1,140 SNVs, 315 were unique-to-GATK, 289 were unique-to-SOAPsnp, and 315 were called by both GATK and SOAP. Among 960 Indels selected, 386 were from the unique-to-GATK Indel set, 387 were from the unique-to-SOAPIndel set, and 187 were from set of Indels overlapping between the two (SOAPIndel and GATK).

PCR primers were designed using the software program Primer 3 (http://sourceforge.net) to produce amplicons of 100 to 200 bp in size containing variants of interest in approximately the center of the amplicon. Primers were obtained in 96-well plate format, 10 μmol/L dilution each (Sigma-Aldrich). All primers were first tested for PCR efficiency using a Hapmap DNA sample

(Coriell Institute for Medical Research, Catalog# NA12864) and DNA polymerase (LongAmp® Taq DNA Polymerase, New England Biolabs). An in-house sample K8101-49685s' genomic DNA was used as template for the validation experiment. After quality-control steps using agarose gel, the product was purified (ExoSAP-IT® reagents, Affymetrix) and pooled. Final PCR products were quantified (Qubit® dsDNA BR Assay Kit, Thermo Fisher Scientific), then library construction for the sequencer platform (MiSeq Personal Sequencer; Illumina) was performed using NEBNext DNA Library Prep Master Mix Set for Illumina protocol. Finally, before being loaded onto the MiSeq machine, the quality and quantity of the sample was verified using the Bioanalyzer (Agilent Technologies) and quantitative PCR (Kapa Biosystems) by staff at the Sequencing shared resources at CSHL.

For Project 2, a total of 1,400 Indels were selected for MiSeq validation, including 1,000 Indels resequenced before the release of GATK v.3.0, and 400 Indels after the release of GATK v.3.0. Among the additional 400 Indels, 215 were covered with more than 1,000 reads in the initial MiSeq data set reported in Project 1. Among the first 1,000 indels we randomly selected, 200 were from the intersection of all pipelines (HapCaller, SOAPindel, and Scaplel), 200 were specific to the respective pipelines, and 200 were ≥30 bp in size that came from the union of all indels detected by these three algorithms. PCR amplicons designed for this project ranged in size from 200 to 350 bp to use the more recent 2x250 bp Illumina MiSeq sequencing protocol. Library construction for the MiSeq Personal Sequencer platform (Illumina) was performed according to the Illumina TruSeq DNA Sample Prep LS protocol (for the initial 1,000 Indels) and TruSeq Nano DNA Sample Preparation Guide (for the additional 185 Indels), omitting the DNA fragmentation step. Finally, the same quality control and quantification experiments were performed before samples were loaded to the MiSeq machine for sequencing.

For Project 3, we randomly selected 200 Indels for validation on an in-house sample K8101-49685s from each of the following categories: (1) Indels called from both WGS and WES data (WGS-WES intersection), (2) WGS-specific Indels, (3) WES-specific Indels. Out of these 600 Indels, 97 were covered with more than 1,000 reads in the previous MiSeq data set reported in Project 2. Hence, we only performed additional Miseq validation on the remaining 503 loci. Sample preparation process was similar to that used in project 2, just omitting the PCR primer testing experiment. Rather, 25 μL PCR reactions were directly set up to amplify each INDEL of interest using K8101-49685s' genomic DNA as template and LongAmp Taq DNA polymerase (New England Biolabs). Library construction was performed following the TruSeq Nano DNA Sample Preparation Guide (Illumina).

NGS data process and analysis

NGS data process and analysis in this study was performed by Jason O'Rawe (project 1, 2), Han Fang (project 2, 3), and Giuseppe Narzisi (project 2). Detailed methods can be found in the original publications (Han Fang, Wu, Narzisi, O'Rawe, Barrón, Rosenbaum, Ronemus, Iossifov, Schatz, & Lyon, 2014; Narzisi et al., 2014; J. O'Rawe et al., 2013). In brief, raw data in project 1 was analyzed using near-default parameters with five different alignment and variant-calling pipelines (SOAP (v.2.21 for SOAPaligner and v.1.03 for SOAPsnp), BWA (v.0.5.9), GATK (v.1.5)/UnifiedGenotyper (v.2.3-9)/HaplotypeCaller (v.2.3-9), SNVer (v. 0.2.1), GNUMAP (v.3.1.0), and SAMtools (v.0.1.18)). And standard data-normalization procedures and canonical genotype-clustering files provided by Illumina were used to process the genotyping microarray signals. For project 2, sequencing reads were aligned using BWA, SAMtools, Picard, and Bowtie2, variants were called using Scalpel (v.0.1.1), GATK HaplotypeCaller (v.2.4.3), and SOAPIndel (v.2.0.1). For analysis in project 3, BWA-mem (v0.7-6a), SAMtools (v0.1.19-

44428cd), Picard (v1.106), Scalpel (v.0.1.1), GATK-UnifiedGenotyper (v3.2-2) (DePristo et al., 2011), BedTools (v2.18.1) (Quinlan & Hall, 2010), lobSTR (v2.04) (Gymrek et al., 2012) were used.

Accessing data

Sequencing data were submitted by Jason O'Rawe (project 1, 2), Han Fang (project 2, 3), and Giuseppe Narzisi (project 2) to the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra/) under project accession number SRP019719 (project 1), SRX265476 & SRX386284 (project 2), and SRX701020 (project 3). Also, all the HiSeq data generated in project 3 from eight SSC samples have been submitted by Ivan Iossifov to the National Database for Autism Research (https://ndar.nih.gov/) under collection 'Wigler SSC autism exome families' (project number 1936).

1.3 Results

Logic of using MiSeq resequencing strategy

Sanger sequencing, the first-generation sequencing method developed by Sanger and colleagues (Sanger et al., 1977) in 1970s, which used dideoxy chain termination, was considered as the gold standard for DNA sequencing for around three decades. And even now, the Sanger method remains the benchmark in the field for accurately determining DNA sequence, with an error rate of less than 1 in 10,000 bases (accuracy > 99.999%). However, for a project that aims to validate thousands of variants identified by a NGS platform, Sanger sequencing becomes unwieldy due to its low throughput nature and high cost ($2,400/million bases vs $0.07/million

bases for HiSeq 2000) (Liu et al., 2012). Hence, we need to develop an efficient and robust validation protocol based on a NGS platform which has a high accuracy.

Oftentimes when a result needs to be validated, an orthogonal testing, which utilizes a different method, can be really useful in some cases. For example, whole genome sequencing can be a very good orthogonal validation method for sequence data generated by exome capture and sequencing (WES), because WGS reads are generated by PCR amplifying the randomly fragmentized DNA, whereas exome capture is based on hybridization of designed probes to the targeted exon regions (Warr et al., 2015). However on the other hand, it also needs to be kept in mind that two sequencing platforms, e.g. Complete Genomics vs Illumina HiSeq, which use completely different sequencing mechanisms and protocols likely will not serve as a good data validation platform for each other, because each of them will have their unique types of systematic error, which can be a cofounding contributor to a low validation rate (Lam et al., 2012).

A recent study compared sequences generated by Ion Torrent PGM, MiSeq and PacBio platforms, and reported all platforms have a near perfect coverage behavior on GC-rich, neutral and moderately AT-rich genomes, however PGM has a profound bias observed upon sequencing the extremely AT-rich genome (no coverage for approximately 30% of the genome). Also, authors discovered a higher false positive rate of SNVs called from PGM compared to Miseq, and a significant higher error rate (over ten times higher) associated with PacBio platform compared to others (Quail et al., 2012).

When putting all this together plus the cost considerations, we ultimately decided Illumina MiSeq platform was our best choice for developing the validation protocol, especially

considering the sequencing data we wanted to validate was generated by Illumina HiSeq2000. Although the observed raw error rate with MiSeq sequencing is a little higher than HiSeq 2000 platform (0.8% vs 0.26%) (Quail et al., 2012), we believed a high depth of coverage with MiSeq data would help compensate for this.

Workflow of MiSeq resequencing

The workflow of a MiSeq resequencing project starts with batch PCR primer design for the targeted genomic loci (Fig 1.2). The task of designing thousands of PCR primers can be achieved through either a high-throughput web tool (O'Halloran, 2015; O'Halloran et al., 2016) or customer developed scripts for automated sequence retrieval, processing and design for batch using commercially available primer design software such as Primer 3 (Untergasser et al., 2012). We designed the primer pairs to product amplicons with the variants of interest located preferably in the center, which helped utilize the maximum power of the paired-end reads from the sequencing platform to increase the confidence during downstream data process and analysis. In total, we designed nearly 4,000 primer pairs for all three validation projects. Primers were ordered in a 96-well plate format with one primer pair per well to facilitate the downstream experiment manipulations using multichannel pipettes or an automated liquid handling workstation.

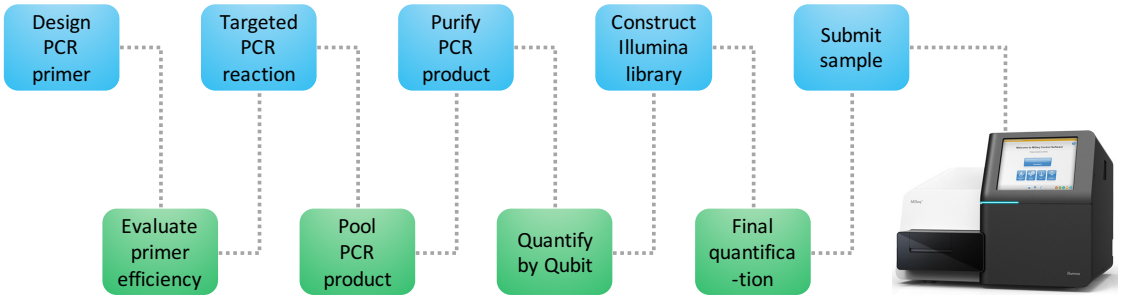
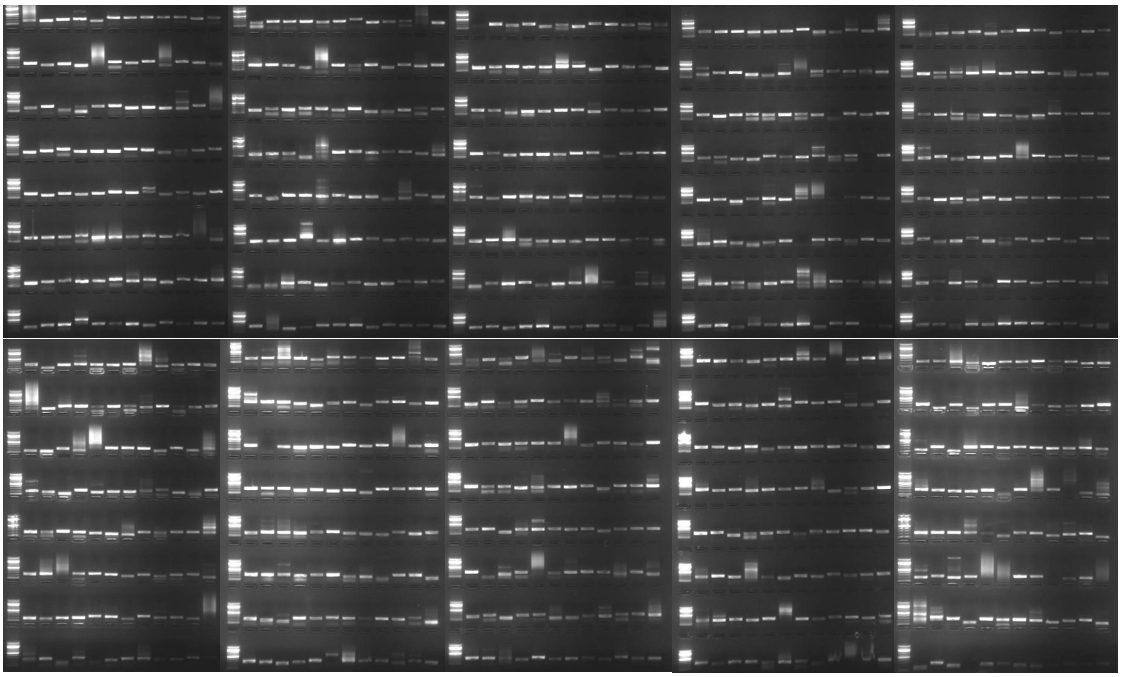
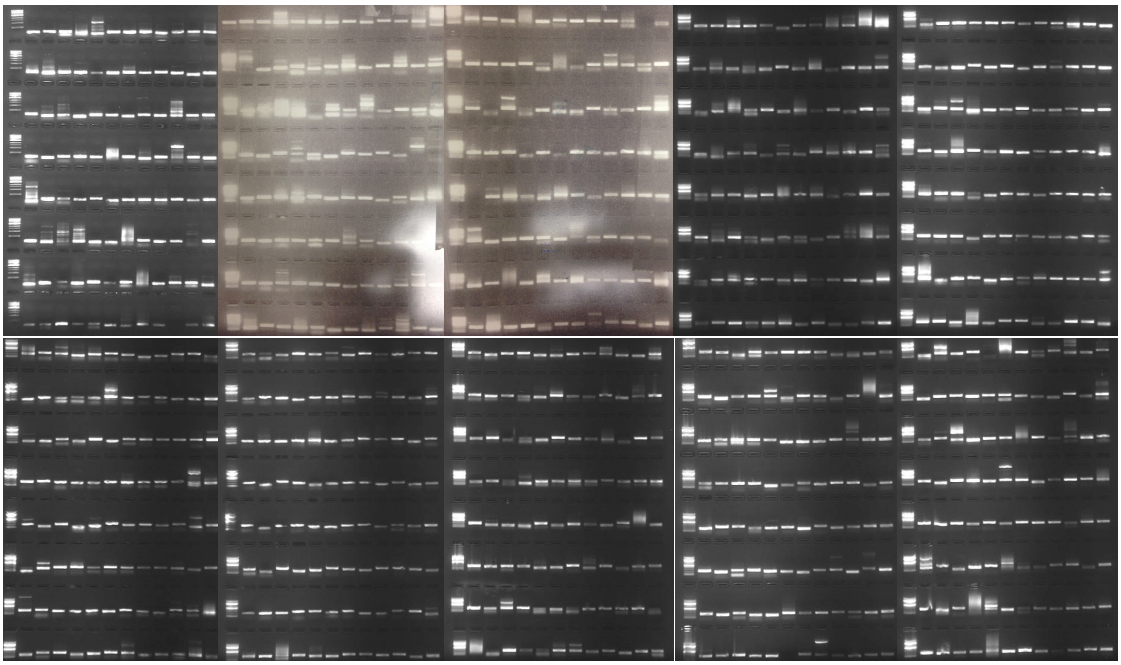


Figure 1.2 General workflow of data generation for targeted resequencing using Illumina MiSeq platform.

We first tested the primers' efficiency when they arrived using a Hapmap sample as template to estimate the success rate of the PCR reactions. It is informative for the downstream analysis especially as some variants to be validated were located within difficult genomic regions for PCR amplification such as a high GC content environment. We observed an overall success rate over 80% at this step (Fig 1.3). This step can be omitted if the primer designing program has been confidently tested before and the success rate of PCR amplification of the same kind was known based on previous experiences.





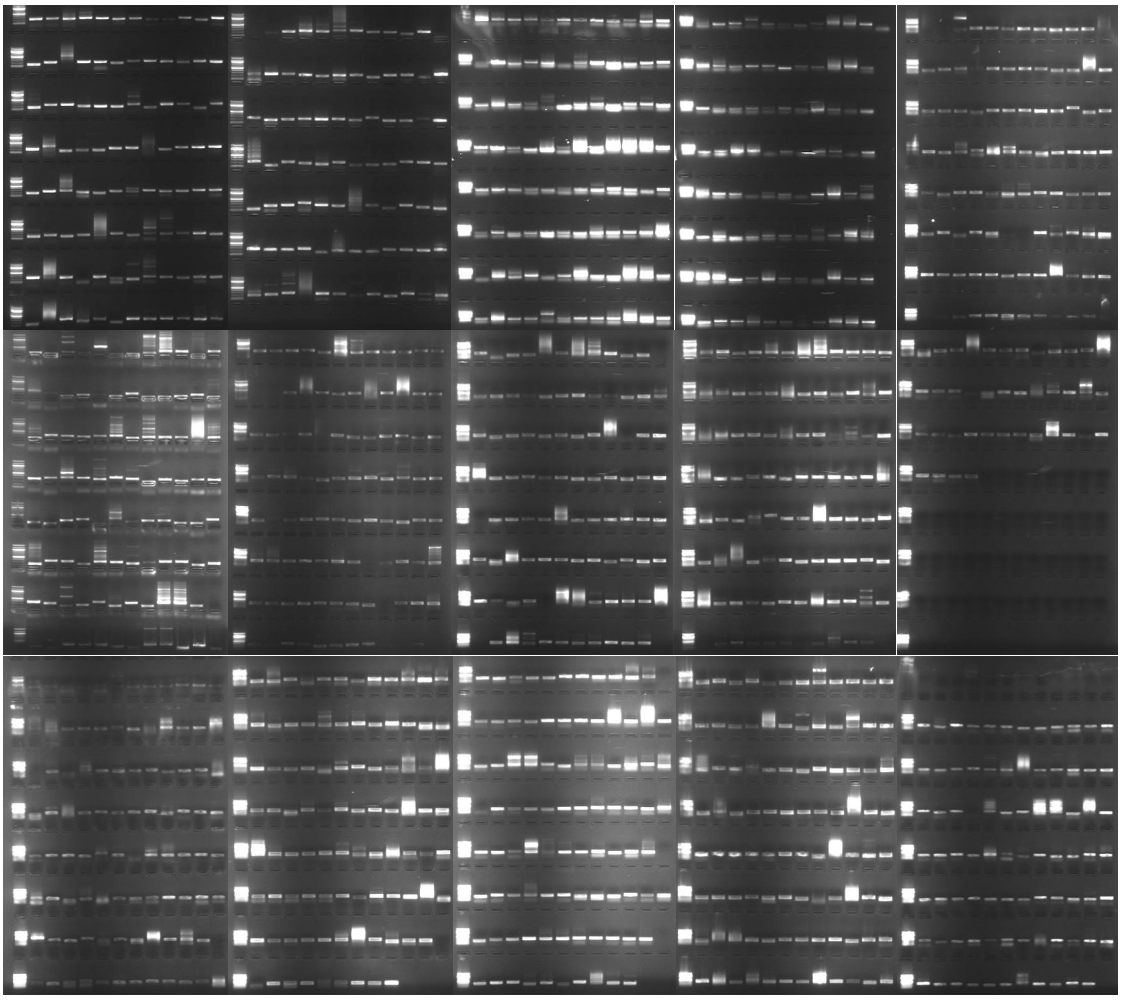
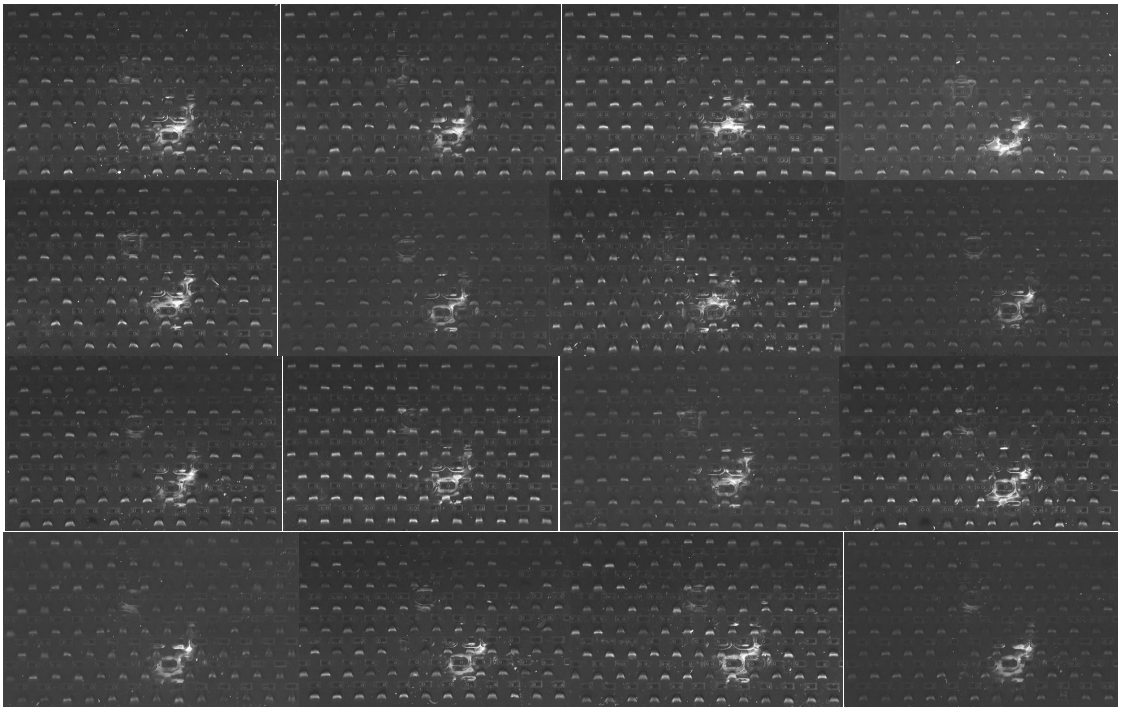
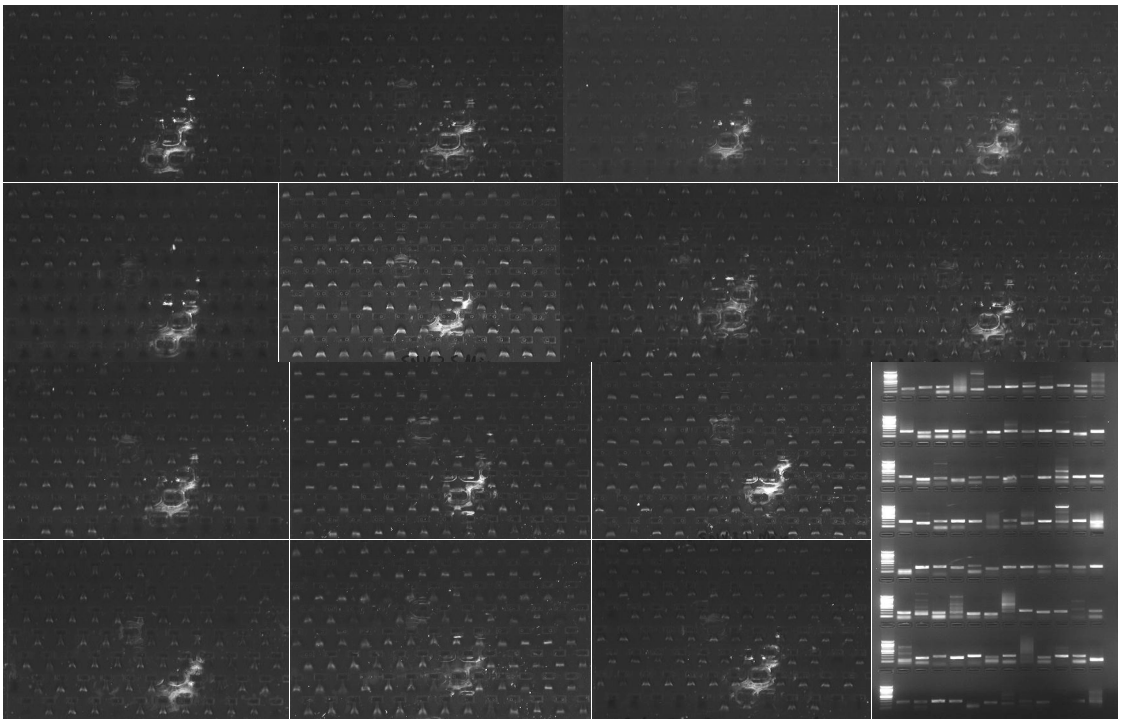


Figure 1.3 Agarose gel images of test PCR reaction during all three validation projects.

Next, we performed the target PCR reactions using the study sample as template, and we evaluated the quality of the amplification by agarose gel again (Fig 1.4). Different genomes could cause mild variations in amplification efficiency, but the overall success rate should be similar as using the test sample. After this, we pooled all the PCR amplicons for cleanup and purification to remove the unused primers and dNTPs (through ExoSAP-IT kit, Affymetrix), as well as enzymes, salts, and other impurities from the sample pool (through QIAquick PCR purification kit, Qiagen) to eliminate the possible interference with the following library construction reactions. Purified samples were quantified by Qubit dsDNA BR assay kit (Thermo Fisher Scientific, # Q32850), followed by standard Illumina Truseq library construction protocols omitting the DNA fragmentation step since our samples were small PCR amplicons rather than genomic DNAs. At the end of the library construction, quality and quantity of the reactions were checked again by bioanalyzer (Fig 1.5) and qPCR. As can see, the average size of the products after each library construction was around the target size (designed amplicon size plus the adaptor length on each side of the amplicon), indicating the successful enrichment of the products of targeted amplicons ligated to the sequencing adaptors. Upon finishing all the quality control, samples were loaded onto the sequencing platform for a MiSeq run.





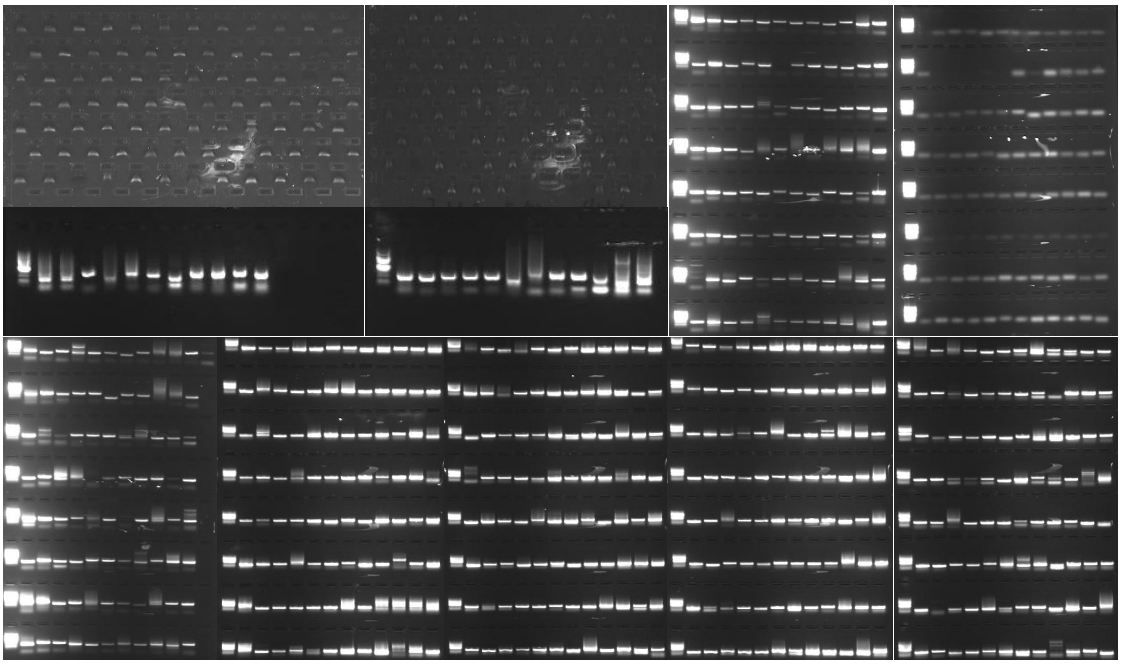


Figure 1.4 Agarose gel images of target PCR reaction during all three validation projects.

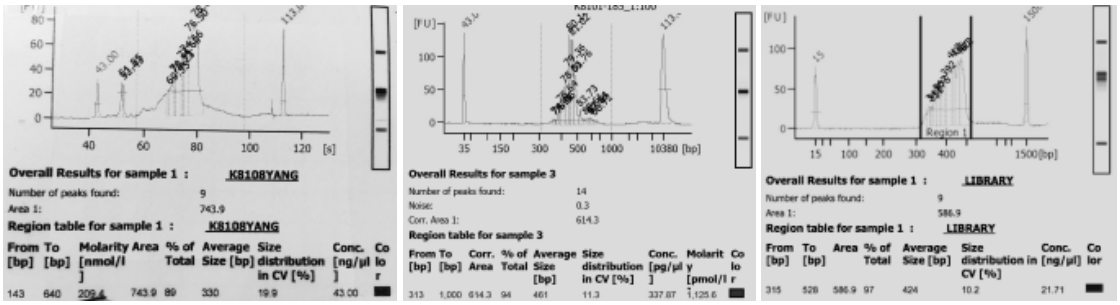


Figure 1.5 Bioanalyzer analysis of the post library construction samples. Average size and quantity of the amplicons after library construction were evaluated by this analysis. As seen, the average size of PCR amplicons after library construction was around 330 bp for project 1, 461 bp for project 2, and 424 bp for project 3.

Data production summary

For project 1, data generated from WGS by CG was with 95% of the exome region covered by 20 reads or more per base, resulting in greater than 88% of the genome covered with a depth of greater than 20 reads per base. Data generated from 15 samples sent out for exon capture and sequencing had a mean sequence coverage of approximately 120× on targeted regions. And for all samples, sequence reads covered more than 80% of the targeted region with a depth of greater than 20 reads per base. For project 3, WGS and WES data from an in-house sample K8101-49685s had a mean coverage of 30× and 110×, respectively.

MiSeq validation rate summary

MiSeq validation data summary was demonstrated in Table 1.2. For project 1, 919 out of 1,140 (81.0%) SNVs targeted for MiSeq validation were successfully amplified and sequenced, with an average read depth of 5,392. For 960 Indels selected for validation, 841 (83.5%) were successfully amplified and sequenced with an average coverage of 4,866. Based on the validation data, 97.1%, 60.2% and 99.1% of the GATK(v.15)-only, SOAPsnp(v1.03)-only and shared SNVs were validated, and in the meanwhile, 54.0%, 44.6% and 78.1% of the GATK-only, SOAP-only and shared Indels were validated (Table 1.3).

Table 1.2 MiSeq validation data summary

		Project 1	Project 2	Project 3
Selected SNVs		1140	N/A	N/A
Selected Indels		960	1400	600
Success rate for amplification and sequencing (> 1,000 reads)	SNVs	81.0%	N/A	N/A
	Indels	83.5%	NK	73.8%
Read length (paired-end)		100 bp	250 bp	250 bp
Mean coverage	SNVs	5,392×	N/A	N/A
	Indels	4,866×	47,018×	55,000×

SNV, single nucleotide variant; Indel, insertion/deletion; N/A, not applicable; NK, not known.

Table 1.3 MiSeq validation rate for SNVs/Indels specific to each pipeline/platform

Project	Type	Tool/Platform	Valid	Invalid	PPV (%)	Valid (≥ 5 bp)	Invalid (≥ 5 bp)	PPV (%) (≥ 5 bp)
Project 1	SNV	GATK	306	9	97.1	N/A	N/A	N/A
		SOAPsnp	174	115	60.2	N/A	N/A	N/A
		Intersection	312	3	99.1	N/A	N/A	N/A
	Indel	GATK	180	156	54.0	NK	NK	NK
		SOAPsnp	148	184	44.6	NK	NK	NK
		Intersection	132	37	78.1	NK	NK	NK
Project 2	Indel	Scalpel	145	43	77.1	13*	1*	92.8
		SOAPindel	101	99	50.5	8*	129*	5.8
		HaplotypeCaller	45	155	22.5	7*	62*	11.3
Project 3	Indel	WGS-specific	122	23	84.1	25	8	75.8
		WES-specific	91	70	56.5	1	0	100
		Intersection	152	8	95	18	0	100

PPV, the positive predictive value computed as #TP/(#TP + #FP), where #TP is the number of true-positive calls and #FP is the number of false-positive calls; N/A, not applicable; NK, not known; WGS, whole genome sequencing; WES, whole exome sequencing. *, Indels ≥ 30 bp. Data was from the original papers (Fang et al., 2014; Narzisi et al., 2014; O'Rawe et al., 2013)

For project 2, our protocol generated high-quality 250 bp reads (paired-end) with an average coverage of 47,018×. Indels detected by all pipelines had a high validation rate, and their sizes followed a log-normal distribution. However, the validation rate varied dramatically for Indels uniquely called by each tool (Table 1.3). Respectively, only 22.5% and 50.5% of the HaplotypeCaller- and SOAPindel-specific indels could be validated even when the less strict position-based approach was used (in which an indel is considered valid if a mutation with the same coordinate exists in the validation data), whereas 77.1% of Scalpel-specific indels were true positive. For the long indels (≥ 30 bp), less than 10% called by SOAPindel (8%) and HaplotypeCaller (7%) passed validation. The new version of GATK (v.3.0) has largely removed the bias toward deletions, but Scalpel still outperformed HaplotypeCaller. Scalpel showed a substantially higher validation rate (86%) for longer indels (> 5 bp) than HaplotypeCaller v.3.0 (27%). Moreover, SOAPindel showed an appreciably higher rate of false positives within microsatellites. When microsatellites were excluded, the performance of SOAPindel and HaplotypeCaller declined, whereas Scalpel's validation rate was only slightly reduced.

For project 3, high quality 250 bp paired-end reads were generated with an average coverage of 55,000 over the selected INDELs. 371 out of the 503 loci selected for validation were covered with over 1,000 reads which were used for the downstream analysis. Among these 371 loci, 160, 145, and 161 loci were from the WGS-WES intersection, WGS-specific, and WES-specific INDELs, respectively. The validation rates of WGS-WES intersection Indels, WGS-specific Indels, and WES-specific ones were 95%, 84%, and 57%, respectively (Table 1.3). Moreover, large Indels (> 5 bp) in the validation set called by both WGS and WES were all valid, while only 76% of those uniquely called by WGS data was validated (Table 1.3). Note that, among the original call set, 9% of the WGS-WES intersection Indels were large ones by our

definition (> 5 bp), and the number of that for WGS-specific Indels is 21%. However, only 1.5% (10) WES specific Indels were large by this definition, and just one got randomly selected for validation. This indicated that the WES data missed most large Indels, which we speculated might be related to an insufficient exome capture caused by a disrupted hybridization to the probes, which would then contribute to a low coverage in those regions.

1.4 Discussion

Before our projects were initiated, there were a few reports discussing the differences in the data generation using different sequencing platforms, which had raised attention to carefully interpret NGS data from different sequencing technologies (Clark et al., 2011; Lam et al., 2012; Quail et al., 2012; Rosenfeld et al., 2012; Suzuki et al., 2011). For example, a group compared the performance of SNV and Indel calls by Illumina and Complete Genomics (CG) platforms and revealed 88.1% concordance in SNV calling but only 26.5% Indels were detected by both platforms (Lam et al., 2012). Meanwhile, the authors noticed that both methods clearly call variants missed by the other technology, especially for indels. Another group reported a comparison study result on inter- and intra-population variances identified by the 1000 genomes project (1KGP) and the Complete Genomics (CG) platform from the same genomes, which revealed a striking degree of difference (> 19%) of SNV calls between the two sets (Rosenfeld, et al., 2012). The authors pointed out that it was likely due to differences in technology, data collection, read-alignment methods, and variant-calling algorithms. However note that all these studies either didn't perform any validation experiment or used validation methods that likely introduced more biases themselves, such as traditional Sanger sequencing (hence, very limited number of variants were selected for validation, < 70 for SNVs or Indels (Lam et al., 2012)),

genotyping arrays (which have the tendency to cover relatively common variants), and target enrichment capture based on Illumina technology (problems with inefficient capture, etc.).

Our group was among the first to establish the targeted resequencing strategy by batch PCR amplifying the loci of interest and sequencing the pooled samples on Illumina MiSeq platform. Based on the validation data summary gathered from all three projects, we were pleased to see that this targeted resequencing strategy has achieved its goal and eliminated the concerns associated with previous existing validation protocols mentioned above. Moreover, the successful application of it to three different projects also proved the robustness of our protocol. The development of a such protocol has facilitated detailed characterization of some sources that can contribute to the uncertainties of NGS data, especially algorithm biases and artifacts, and also enhanced our understanding of the behaviors of variant calls during NGS data analysis, especially Indel calling, which is in general much harder to call and call accurately than SNVs.

The three projects discussed in the results were initiated independently to address different technique aspects of NGS, with project 1 aiming to gain a greater understanding of the variability introduced into SNVs and Indels discovery when utilizing different bioinformatics pipelines and different sequencing platforms (O'Rawe et al., 2013), project 2 was to investigate the performance of a newly developed, open-source, microassembly Indel calling algorithm Scalpel (Fang et al., 2016; Narzisi et al., 2014), by compared it to other similar software, and project 3 was to understand behaviors of the Indel calls from the WGS and WES data, and to characterize the sources of Indel calling errors in general (Fang et al., 2014). Their observations together provide some of the first evidence of the low concordance of variant-calling pipelines and variant detection biases in WGS and WES.

We observed significant discrepancy in SNV and Indel calling between many of the variant calling pipelines available when analyzing the same sets of sequence data, to be specific, with 57.4% concordance for SNV (between BWA-GATK, SOAP-Align-SOAPsnp, BWA-SNVer, BWA-SAMtools, and GNUMAP-GNUMAP) and 26.8% for Indel (between GATK, SOAPindel, and SAMtools) (project 1) (J. O'Rawe et al., 2013), and about 37% concordance among Indel calls made by Scalpel, SOAPindel, and HapCaller (project 2) (Narzisi et al., 2014), which demonstrated fundamental, methodological variation between these commonly used bioinformatics pipelines. Moreover, we also demonstrated the low concordance (53%) of Indel detection between WGS and WES, with WGS data uniquely identifying 10.8-fold more high-quality Indels and WES missing many large Indels (Fang et al., 2014). The same study (project 3) also pointed out that homopolymer A/T Indels are a major source of low-quality Indel calls, which are highly enriched in the WES data, and a PCR amplification based library construction could be a major source of low quality poly-A/T Indels (Fang et al., 2014). But at the same time, in spite of these inter-methodological variations we observed, our studies also reveal overall very high validation rates of variants called in the shared data sets either between all pipelines or between different sequencing platforms, indicating that sequencing and analyzing samples with multiple platforms and methodologies might be helpful to attain a high accuracy personal genome.

After we published our work, more and more groups joined our effort to investigate biases and sources of error in NGS and bioinformatics methods (Cornish & Guda, 2015; Highnam et al., 2015; Hwang et al., 2015; Liu et al., 2013; Rieber et al., 2013; Yu & Sun, 2013; Zook et al., 2014). Hence, the need for adopting a solid validation protocol is increasing, especially with the promotion of precision medicine in both research and clinical settings. I

believe in the foreseeable future, with the continuous cost drop with the WGS technologies, more and more researchers will feel comfortable to switch to a WGS-based validation protocol (instead of e.g. array-based validation) like ours for their projects.

Lastly, I wish to advocate for bench scientists to play a more active role in sequencing projects in order to facilitate the quality advancement of genomics data. Many researchers such as myself who are not trained to do script-based bioinformatics analysis of NGS data might feel our responsibility for generating a reliable individual genome only stays with the data generation stage to assure the quality of our bench work. However, with the development of more and more commercialized, user-friendly NGS data analysis software, we can involve ourselves in the downstream NGS quality control. For example, we can visualize BAM files, which are generated after short reads are aligned to the reference genome (i.e. mapping), by using software like Golden Helix GenomeBrowse (available from http://www.goldenhelix.com) to easily assess the quality of the reads. Fig 1.6 shows an example of high quality reads identified in the BAM file which gave me high confidence to call variants within this region. As seen, the coverage of reads from both directions was quite high (near 4,000×), and the distribution of them was uniform along the full length. Examples of poor quality reads are shown in Fig 1.7- Fig 1.9, where reads were extremely biased toward one strand (Fig 1.7), or with very few full length reads present in either strand (Fig 1.8), or with problematic ends (Fig 1.9). The latter observation was also validated by the VCF file (Variant Call Format), shown in the lower panel of Fig 1.9, and revealed the poor quality of the reads at the end. These quality checks are relatively straightforward yet very helpful, and it should be implemented by both bioinformaticians and other researchers that are involved in a sequencing project.

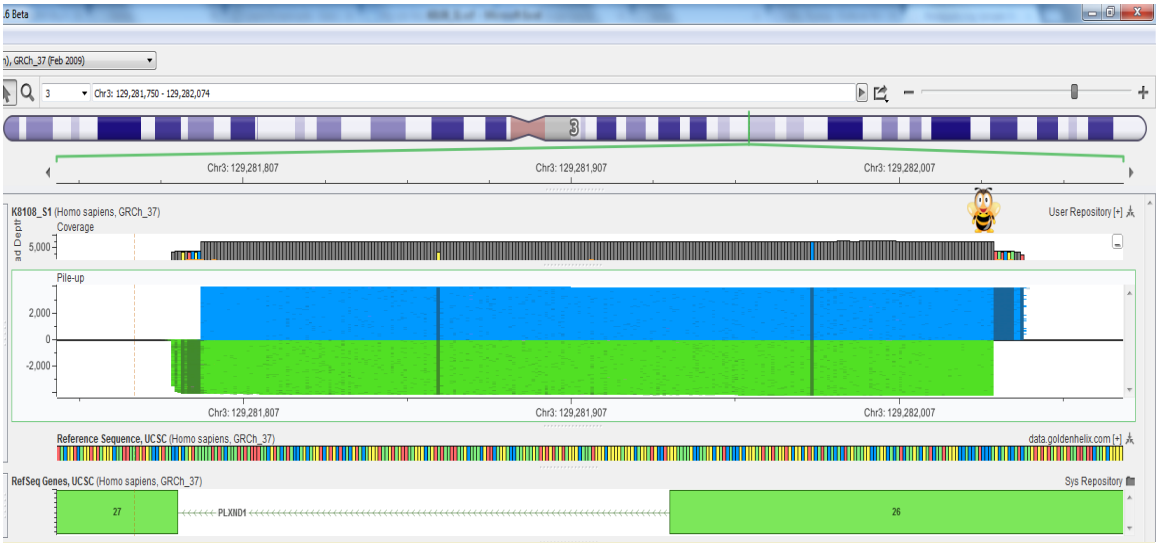


Figure 1.6 An example of high quality reads from both strands (BAM file). Blue and green colors represent reads from different directions. Screenshots are from Golden Helix GenomeBrowse ® visualization tool v2.x by Golden Helix, Inc.

Figure 1.7 Two examples of reads with strand bias (BAM file). Blue and green colors represent reads from different directions. Screenshots are from Golden Helix GenomeBrowse ® visualization tool v2.x by Golden Helix, Inc.

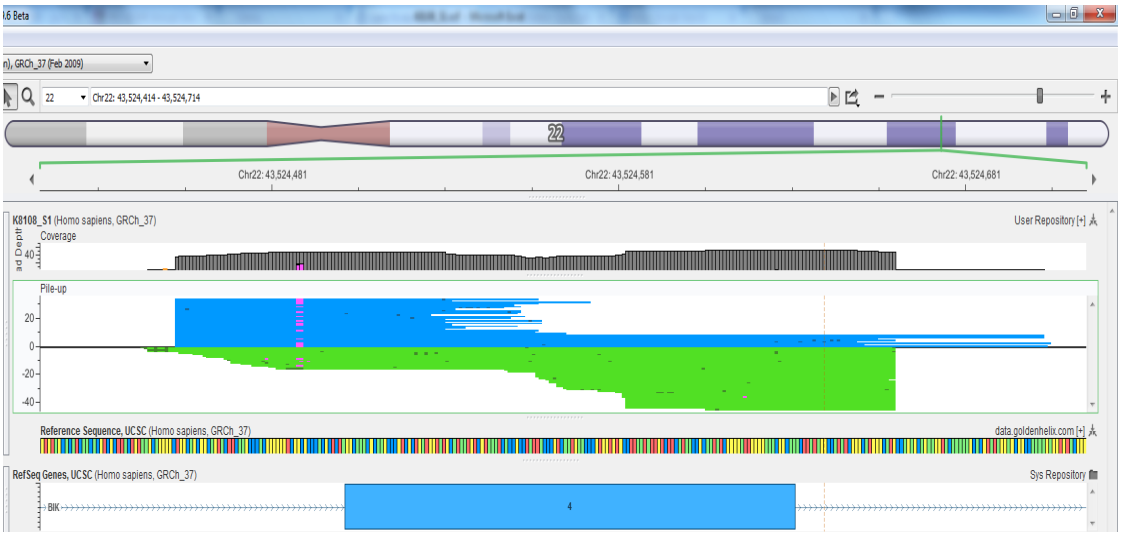
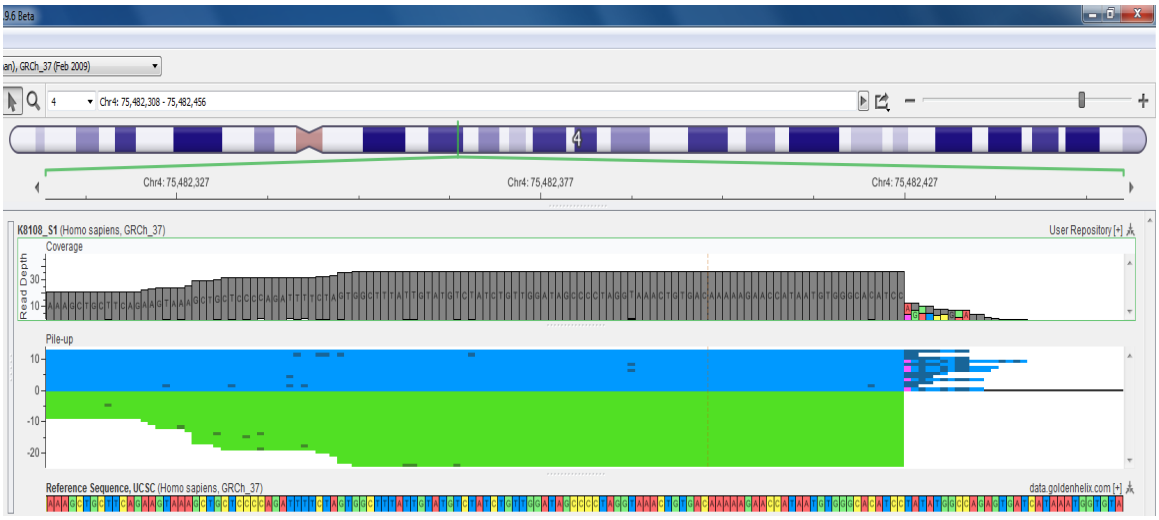


Figure 1.8 An example of reads with interesting patterns (BAM file). Blue and green colors represent reads from different directions. Screenshots are from Golden Helix GenomeBrowse ® visualization tool v2.x by Golden Helix, Inc.



chr4	######	.	CT	C	39	PASS	AC=2;AF=1.	GT:AD:DP:G	1/1:8,5:13:39.07:471,39,0:0.385
chr4	######	.	A	G	99	PASS	AC=1;AF=0.	GT:AD:DP:G	0/1:5,8:13:99:177,0,155:0.615
chr4	######	.	T	A	15	q30	AC=2;AF=1.	GT:AD:DP:G	1/1:0,6:11:15.03:164,15,0:1.000
chr4	######	.	A	T	94	PASS	AC=1;AF=0.	GT:AD:DP:G	0/1:5,6:11:94.20:94,0,156:0.545
chr4	######	.	T	C	63	PASS	AC=1;AF=0.	GT:AD:DP:G	0/1:5,4:9:63.90:64,0,138:0.444
chr4	######	.	G	C	88	PASS	AC=1;AF=0.	GT:AD:DP:G	0/1:4,5:9:88.56:136,0,89:0.556
chr4	######	.	C	G	12	q30	AC=2;AF=1.	GT:AD:DP:G	1/1:0,5:8:12.03:145,12,0:1.000
chr4	######	.	C	A	21	q30	AC=2;AF=1.	GT:AD:DP:G	1/1:0,8:8:21.05:244,21,0:1.000

Figure 1.9 An example of reads with end problems (BAM file). Blue and green colors in the upper panel image represent reads from different directions. Screenshots are from Golden Helix GenomeBrowse ® visualization tool v2.x by Golden Helix, Inc. Variant calls at the end of the reads from VCF file are shown in the lower panel.

1.5 Concluding Remarks

The use of next-generation sequencing technologies as a molecular diagnostic tool in clinical genomics to help medical decision making requires the accurate and consistent identification of personal genome variants. However, the lack of guidelines for using multiple NGS sequencing platforms and bioinformatics tools, and the lack of characterizations of the hidden sources of biases and errors during variant calling have diminished the credibility of personal genome information being used in a clinical setting. Prior to our study, several groups had started to explore the differences of the sequencing data generated by various sequencing platforms; however, the methodological variation between commonly used variant-calling pipelines had yet to be characterized. And in order to do that, a robust, high-throughput, and cost-effective validation protocol needed to be in place to help identify true calls.

We developed of a targeted resequencing protocol based on an approach of PCR amplification of genomic DNA regions containing selected variants followed by pooled MiSeq sequencing. We demonstrated the successful application of this protocol to three independent projects aiming to characterize the comparability and differences of multiple sequencing platforms and variant-calling tools, as well as to explore the signatures of biases and errors associated with variant calling. These applications demonstrated the robustness of our protocol, which can generate high quality validation data within a short amount of time. We encourage researchers to use our validation methods and/or data for evaluating their call sets as well.

In these three applications, we show that the concordance of multiple variant-calling pipelines when applied to the same sequence data set is surprisingly low for SNVs, and even lower for Indels, which demonstrated fundamental differences during algorithm design. And our

validation data reveal that each single existing exon-capture, NGS platform, or variant-calling pipeline is likely to miss some true functional rare variants. However, despite the presence of such differences, we are able to validate most of the variants that shared among all pipelines even under lax parameterization. Therefore, considering current sequencing prices, we suggest the utilization of the total list of variants derived from multiple variant-calling pipelines to be a feasible solution for reducing false-negatives in a discovery setting. Furthermore, we also illustrate the low concordance of Indel detection by WGS and WES, and reasons for that included the capture inefficiency and more coverage biases with WES, which might also cause WES to miss many large Indels as we saw. For this, we recommend to use WGS, higher coverage (> 60×), and micro-assembly-based algorithms such as Scalpel for studies focusing on Indel calls. Lastly, we pinpoint that homopolymer A/T Indels which are highly enriched in WES data is a major source of low-quality Indels. Because of that, we recommend to control for homopolymer false Indel calls with a more stringent filtering criteria. This will be an useful strategy especially for population-scale sequencing projects, because the expense of performing large scales of experimental validation can be overwhelming.

Our findings illustrate the quality uncertainties of current NGS technologies introduced by multiple sequencing platforms and variant-calling pipelines, and it will help inform researchers who are seeking standardization of performance benchmarking of the variant-calling pipelines in the realm of genomic medicine (Zook et al., 2014).

Chapter 2

Accelerating discovery of the genetic basis of disease via clinical genomics

2.1 Introduction

Clinical genomics, which uses the personal genome information generated by the rapid evolving NGS to guide disease diagnosis and patient care, is revolutionizing modern medicine and biomedical research (Biesecker & Green, 2014; Bodian et al., 2016; Delaney et al., 2016; Gagan & Allen, 2015; Willig et al., 2015). It has enhanced the molecular diagnostic capability for many known rare and inherited diseases as early as the neonatal stage (Baker et al., 2016;

Futema et al., 2012; Prada et al., 2014), and also facilitated the discovery of many novel variants linked to human health and disease (Hoischen et al., 2010; Johnson et al., 2010; Kan et al., 2013; Renton et al., 2011; Rope et al., 2011; Wright et al., 2015). However, despite much success using NGS-based techniques to identify disease-contributory variants, there are still many practical issues and controversies in applying exome- or genome-wide NGS-based techniques in a clinical setting (besides the analytic validity issues that I already discussed in the previous chapter), such as ethics issues (Char, 2015; Lyon, 2012; Lyon & Segal, 2013; Solomon, 2015). Meanwhile on the other hand, researchers who have an intense interest to study clinically challenging cases using NGS feel the hurdles on a daily basis to reach to the patients and families and their physicians for collecting clinical data and samples to facilitate their scientific discovery.

Accessing phenotypic data

Clinical data, including patient personal information (e.g. genetic background, pedigree information), symptoms, and medical record, is of great importance to guide the design of a sequencing project aiming to find the disease-contributory variants. And it can also facilitate the prioritization of the variants and offer clues for the underlying molecular pathways (Goh et al., 2007; Köhler et al., 2008). But unfortunately, the significance of collecting and reporting clinical data has been overlooked by researchers in reality. The current publication guidelines for reporting disease-relevant variant discoveries does not require a detailed description or annotation of the clinical data, which diminishes the motivation for researchers to reach out for such information in the first place, especially for those who are not equipped clinical training. Practical settings, regulations and policies, which were established to protect patients' rights ("Federal Policy for the Protection of Human Subjects ('Common Rule')," 1991, "The HIPAA Privacy Rule," 2000), have now become a major hurdle for prohibiting such valuable

information from being shared between patients, families, physicians, and researchers. With the progression of precision medicine, and a more active role patients are playing in research, such hurdles need to be lifted to facilitate more efficient communication and information exchange between various parties. In fact, policy makers including FDA, have started to realize the current challenges and hence have issued guidelines to help pave the way for precision medicine (Food and Drug Administration, 2013). However, the difficulties are expected to remain for the foreseeable future due to the scale of things (involving issues mentioned above) needing to be reformed.

Standardized vocabulary use

Another practical challenge for utilizing clinical information to its full power even without accessibility issues is the lack of standardized vocabulary use in medical documentation (Allanson et al., 2009; Biesecker, 2005), especially in medical genetics departments, where many patients and families are recruited for research. This has also hampered the use of computational tools to analyze clinical data such as the application of artificial intelligence in precision medicine like IBM Watson (Rhrissorrakrai et al., 2016). Many tools have thus been developed to address such issue, and one of them is Human Phenotype Ontology (HPO), which provides a standardized vocabulary of phenotypic abnormalities to describe presentations of human pathologies (Kohler et al., 2017; Robinson & Mundlos, 2010; Robinson et al., 2008).

Throughout my PhD study, I have had the privilege to experience first-hand both the power of clinical genomics and many of these related issues. In this chapter, I show two examples to illustrate the opportunities and challenges we are currently facing in using NGS-

based clinical genomics to accelerate the discovery of the genetic basis of clinically challenging cases.

2.2 Materials and Methods

Enrollment of research participants and consent

This portion of the project was performed by others. For project 1, the collection and the analysis of DNA was conducted by the Utah Foundation for Biomedical Research, as approved by the Institutional Review Board (IRB) (Plantation, Florida). Written informed consent was also obtained from all study participants, and research was carried out in compliance with the Helsinki Declaration. For project 2, the collection and analysis of the DNA used in this study was conducted by the Utah Foundation for Biomedical Research, Protocol #100, approved by Ethical & Independent Review Services, Inc. Written informed consent (including parental consent in children under the age of 18) to participate in research including sample collection and to publish their personal and clinical details relevant to the study including parents' ethnicity, and video and facial photography, was obtained from all participants in the study. Research was carried out in compliance with the Federal Policy for the Protection of Human Subjects 45C.F.R.46.

Clinical evaluation/diagnostics

All tests were prescribed by physicians and performed in the clinic, and I was responsible for systematically reviewing and organizing medical records.

For project 1, a broad range of clinical diagnostic testing was performed on both affected male siblings, including karyotyping, a custom high resolution X-chromosome CGH array (720K

Chromosome X Specific Array from Roche NimbleGen, Inc. USA), subtelomeric FISH study, methylation study for Angelman syndrome, XNP sequencing for ATRX, and fragile X DNA testing, serum amino acid levels, urine organic acids levels, sweat chloride levels, plasma carnitine profile, immunoglobulin levels, urine mucopolysaccharidosis (MPS) screening and thyroid profile examination. Cranial ultrasound was performed and brain imagery was obtained using magnetic resonance imaging (MRI) and computed tomography scanning techniques. Images of the spine were also obtained using MRI. Moreover, cerebrospinal fluid (CSF) was collected from the elder sibling, and neurotransmitter metabolites, tetrahydrobiopterin (BH4) and neopterin (N) profile were screened. In addition, X-chromosome skewing assays were performed on the female participants in the study.

For project 2, the family was interviewed by Gholson Lyon, a board-certified child, adolescent and adult psychiatrist. Medical records were obtained and reviewed, in conjunction with further interviews with the family. The interviews were videotaped and later reviewed to facilitate further diagnostic efforts. Various clinical diagnostic tests were performed by several clinicians on K10031-10133, including tilt table test, brain MRI, ultrasound of the kidneys and chest X-ray. In addition, her cholesterol level, thyroid profile, urine vanillylmandelic acid (VMA), catecholamines panel (urine-free), basic metabolic panel (BMP), and epinephrine and norepinephrine levels were also screened. For K10031-10232, the following diagnostic evaluations were performed: multiple sleep latency test (MSLT), autism diagnostic observation system (ADOS) - module 2, the Childhood Autism Rating Scale (CARS), Behavior Assessment System for Children (BASC), Intelligence Quotient (IQ), Abnormal Involuntary Movement Scale (AIMS), as well as electrocardiogram (EKG), polysomnographic report, and echocardiogram.

Whole genome sequencing and analysis

My role here was to prepare samples for sequencing, and I also performed library construction for some samples. For project 1, whole genome sequencing was used for 10 members of one family. The parents and two affected children were sequenced with two sequencing platforms, Complete Genomics (CG) and Illumina HiSeq 2000, and 6 other related members were sequenced using the Illumina HiSeq 2000 platform. Illumina sequencing libraries were generated from 100 ng of genomic DNA using the Illumina TruSeq Nano LT kit according to manufacturer recommendations. For project 2, whole genome sequencing was used to genotype 9 members (K10031-10143, 10144, 10145, 10235, 10133, 10138, 10231, 10232, 10233) from the extended pedigree. All libraries were constructed with PCR amplification, and sequenced on Illumina HiSeq2000 with an average paired-end read length of 100 bp. Microarray data for the same samples were generated with the Illumina Omni 2.5 microarray at the Center for Applied Genomics Core of the Children's Hospital of Philidephia.

The sequencing data analysis was carried out by scripts written by Jason O'Rawe (for project 1) and Han Fang (for project 2) using multiple analytic pipelines, and detailed descriptions and data production summary can be found in the original publication (Fang et al., 2017; O'Rawe et al., 2015). I was deeply involved in the study design for these projects and also contributed greatly during the variant prioritization process.

Sanger sequencing

Polymerase chain reaction (PCR) primers for alleles of interest were designed to produce amplicons of 680-850bp in length using Primer3 (http://primer3.sourceforge.net). Primers were obtained from Sigma-Aldrich, and tested for PCR efficiency with an in-house DNA sample using

the indicated DNA Polymerase systems. The optimized PCR reactions were then carried out using DNA extracted from samples of interest as templates. PCR products were visually inspected for amplification efficiency using agarose gel electrophoresis and were purified using the QIAquick PCR Purification Kit (QIAGEN). Purified products were then diluted to 5-10 ng/μL in water for use with the ABI 3700 sequencer. The resulting ∗.ab1 sequence files were loaded into the CodonCode Aligner V5.1.2 for analysis. All sequence traces were manually reviewed to ensure the reliability of the genotype calls.

PCR Primers used in the study: (5'-3', forward/reverse):

ASB12: ctacaacaactgttctcccgt/agccctaatgctctatcctt
ZNF41: gtactgaatgtgtaatgggct/tgagttttctgatgtgtggtgag
PION: gacggcctacttctgtggtgac/ggctgggtgtgggtggca
TAF1: tggcttctggttgatcagtg/aaacgccatcaaaattcctg
POU4F1: caagagccatcctttcaagc/tgtgcatatgcgggtgag

Availability of data and materials

For project 1, all sequencing data have been deposited (I was not involved) to the Sequence Read Archive (SRA) under BioProject ID PRJNA301337. The study accession number is SRA: SRP065848, and the sample accession numbers are SRA: SRS1153874, SRS1153892, SRS1153893, SRS1153896, SRS1153897, SRS1153898, SRS1153899, SRS1153900, SRS1153901, SRS1153902. For project 2, all of the sequence reads can be downloaded under project accession number [SRP058003] from the Sequence Read Archive. Administrative permission was received from the Utah Foundation for Biomedical Research to access the medical records reviewed in this study.

2.3 Results

Project 1. The TAF1 story

Clinical evaluations and phenotypic presentation

All clinical information was collected by various physicians, but was systematically analyzed and organized by myself.

In this study, a Caucasian family was recruited from Utah consisting of three generations of 10 individuals (Fig 2.1). The affected probands were two brothers, ages 12- and 14-years-old respectively, with severe intellectual disability (ID), autistic behaviors, anxiety, attention deficit hyperactivity disorder, and very distinctive facial features (Fig 2.2, 2.3, Table 2.1). Among the facial features are a broad, upturned nose, sagging cheeks, downward sloping palpebral fissures, prominent periorbital ridges, deep-set eyes, relative hypertelorism, thin upper lip, a high-arched palate, prominent ears with thickened helices, and a pointed chin. Other shared phenotypic symptoms include strabismus (exotropia), blocked tear ducts, microcephaly, mild ventriculomegaly, deficiency of the septum pellucidum, hypoplasia of the corpus callosum, low cerebral white matter volume, oculomotor dysfunction, frequent otitis media with effusion, hearing impairments (mixed conductive/sensorineural), oral motor dysphagia, kyphosis, a peculiar gluteal crease with sacral caudal remnant (without any spinal abnormalities), dysplastic toenails, hyperextensible joints (especially fingers and wrists), spasticity, ataxia, gait abnormalities, growth retardation and global develop-mental delays, especially in the areas of gross motor and verbal expression. The younger of the two affected siblings also suffers from frequent episodes of contact dermatitis and eczema, scoliosis, sleep-wake dysregulation, as well

as asthma, although he no longer requires medication for the latter. The elder brother, on the other hand, has diplegia, and has received Botulinum Toxin (Botox) therapy for his lower-extremity spasticity for six years. A review of systems (ROS) questionnaire revealed no other obvious, shared or otherwise, symptoms or malformations. See Table 2.1 for a comprehensive list of clinical features that are present in these probands. In an effort to help standardize phenotype reporting, we used Human Phenotype Ontology (HPO) annotation.

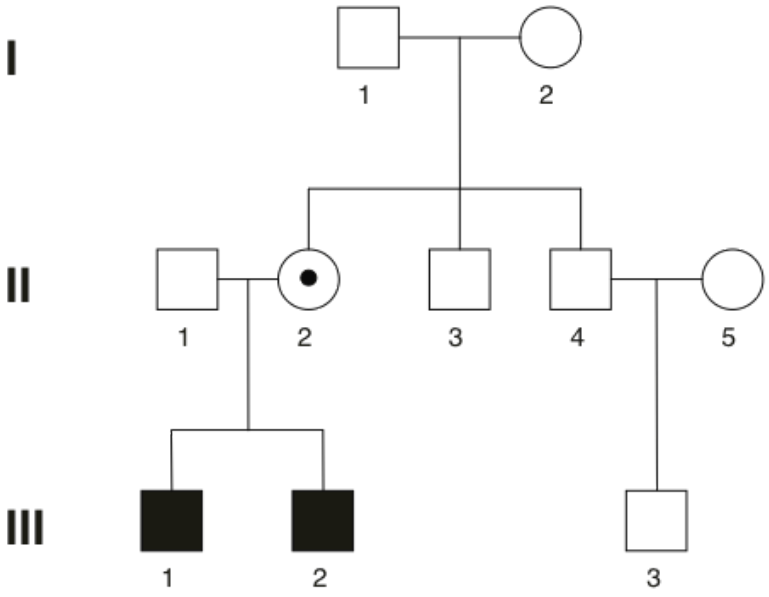


Figure 2.1 Pedigree of the studied family in project 1 (the TAF1 story). This Caucasian family was recruited from Utah, USA, consisting of three generations and ten individuals. Two affected probands III-1 and III-2 are indicated with dark squares. The black dot indicates the maternal carrier.

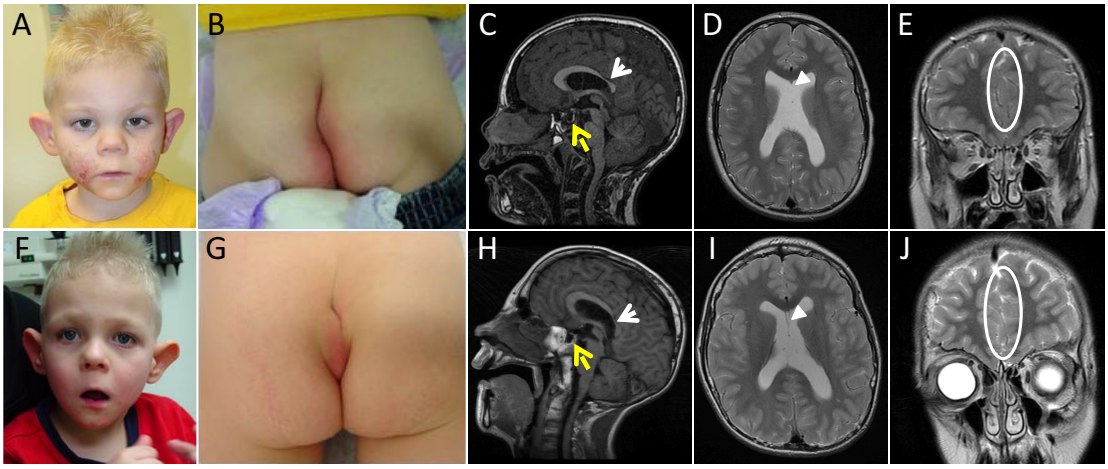


Figure 2.2 Facial features, peculiar gluteal crease and MRI images of probands discussed in project 1. The affected brothers, III-1 at the age of 5 (F) and III-2 at the age of 3.5 (A), have distinctive shared facial features, including a broad, upturned nose, sagging cheeks, downward sloping palpebral fissures, prominent periorbital ridges, deep-set eyes, relative hypertelorism, thin upper lip, a high-arched palate, prominent ears with thickened helices, and a pointed chin. Both affected males have a sacral dimple and an abnormal gluteal crease (B, G). Sagittal T1 (C, H), axial T2 (D, I), and coronal T2 (E, J) weighted brain MRIs of the two brothers from the Utah family (younger brother C-E, older brother H-J) demonstrated a remarkably similar constellation of abnormalities, which includes hypoplasia of the isthmus and splenium of the corpus callosum (white arrows), deficiency of the septum pellucidum in both brothers (arrowheads), as well as findings associated with septooptic dysplasia, including underdeveloped pituitary glands for age (yellow arrows), deficiency of the anterior falx with mild hemispheric interdigitiation (circles), and question of small olfactory bulbs despite fully formed olfactory sulci. Clinical data were collected by Dr. Lyon, and organized and analyzed by myself. MRI images were analyzed by Dr. Edward Yang at Boston Children's hospital. Some images were also used in our paper (O'Rawe et al., 2015).

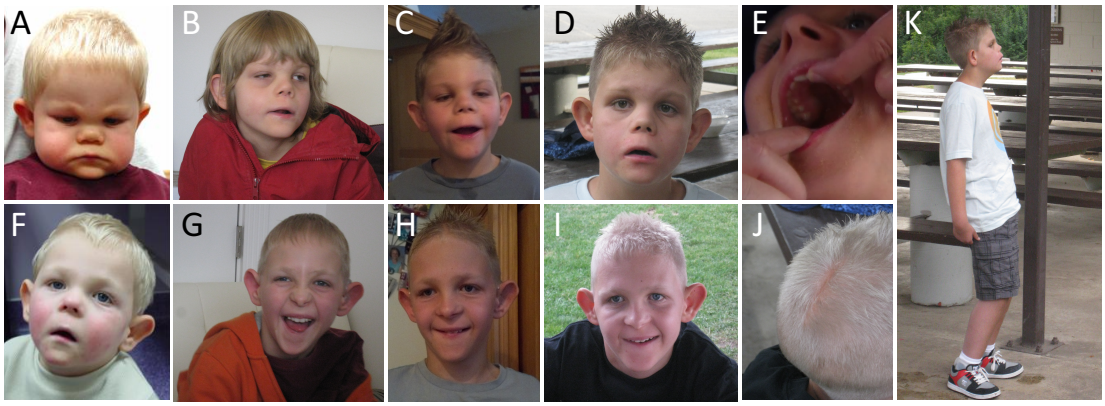


Figure 2.3 Additional phenotypic images of probands discussed in project 1. The younger brother at the ages of 19 months (A), 9 (B), 10 (C), and 12 (D) years old, and the elder brother at the ages of 3 (F), 11 (G), 12 (H), and 14 (I) years old share distinctive facial features, including a high-arched palate (E). The elder boy has a notable scar on his head from the surgery of treating his birth defect of aplasia cutis congenital (J). The younger boy at the age of 12 shows a unique standing posture with protruding abdomen and bent knees (K). Clinical data were collected by Dr. Lyon, and organized and analyzed by myself. This figure I made were also used in our paper (O'Rawe et al., 2015).

Table 2.1 Clinical features of probands in the TAF1 story

Systems	Features	Human Phenotype Ontology (HPO) #	Family 1 (Utah)	
			III-1	III-2
			(elder)	(younger)
General Info	Gender		male	male
	Age		15	13
	Genetic Background		Caucasian	
Genetic Studies	Karyotype		46,X,inv(Y)(p11.2q11.2)	
	TAF1 Mutation (hg19)		chrX:70621541 T>C, p.I1337T	
	Inheritance Pattern		inherited from the mother (De Novo)	
Gestation	Complications		placenta deterioration	
	Term (Weeks)		40	37
Birth	Caesarian Section	HP:0011410	+	
	Weight (Centile)		2.21 kg	1.76 kg
	Length (Centile)		NK	NK
	Head Circumference (Centile)		NK	NK
	Apgar Scores		NK	NK
Perinatal Course	Complications		NK	neonatal jaundice, poor feeding (HP:0011968)
Growth	Intrauterine Growth Retardation	HP:0001511	+	
	Postnatal Growth Retardation	HP:0008897	+	
Neurobehavioral/Development	Delayed Gross Motor Development	HP:0002194	+	
	Delayed Speech and Language Development	HP:0000750	+	
	Oral-pharyngeal Dysphagia	HP:0200136	+	
Craniofacial	Prominent Supraorbital Ridges	HP:0000336	+	
	Downslanted Palpebral Fissures	HP:0000494	+	
	Deeply-set Eye	HP:0000490	+	
	Prominent Forehead	HP:0011220	+	
	Sagging Cheeks		+	
	Long Philtrum	HP:0000343	+	
	Low-set Ears	HP:0000369	+	
	Protruding Ear	HP:0000411	+	
	Thickened Helices	(Large earlobe) HP:0009748	+	
	Long Face	HP:0000276	+	
	High Arched Palate	(High palate) HP:0000218	+	
	Microretrognathia	HP:0000308	+	
	Thin Upper Lip	(Thin upper lip vermilion) HP:0000219	+	
	Pointed Chin	HP:0000307	+	
	Broad Upturned Nose	(Anteverted nares) HP:0000463	-	
	Bulbous Nasal Tip	(Bulbous nose) HP:0000414	-	
	Depigmentation	(depigmentation/hyperpigmentation of skin) HP:0007483	+	
Skin	Hypertelorism	HP:0000316	+ (on the scalp)	-
	Aplasia Cutis Congenita	HP:0001057	+	
	Hirsutism	HP:0001007	-	-
	Frequent Dermatitis & Eczema	(Eczema) HP:0000964	-	+

Ear, Nose, Mouth, and Throat	Toenail Dysplasia	HP:0100797	+	+
	Hearing Impairment	HP:0000365	+ (mixed, HP:0000410)	
Eyes	Chromic Otitis Media	HP:0000389	+	
	Strabismus	HP:0000486	+	
	Nasolacrimal Duct Obstruction	HP:0000579	+	
Gastrointestinal	Oculomotor Dysfuntion	(Abnormality of eye movement) HP:0000496	+	
	Constipation	HP:0002019	-	+
Neurological	Gastroesophageal Reflux	HP:0002020	-	+
	Microcephaly	HP:0000252	+	
	Mild Ventriculomegaly	HP:0002119	+	
	Hypoplasia of the Corpus Callosum	HP:0002079	+ (Posterior>Anterior)	
	Hypoplasia of the Cerebellar Vermis	HP:0001320	+	
	Deficiency of the Septum Pellucidum	(Abnormality of the septum pellucidum) HP:0007375	+	
	Deficiency of the Falx Cerebri		+	
	Hypotonia	(Generalized hypotonia) HP:0001290	+	
	Gait Abnormalities	(Gait disturbance) HP:0001288	+	
	Diplegia	(Spastic diplegia) HP:0001264	+	-
Musculoskeletal	Ataxia	HP:0001251	+	
	Sleep-wake Cycle Disturbance	HP:0006979	-	+
	Osteopenia	HP:0000938	+	NK
	Unusual Gluteal Crease with Sacral Caudal Remnant/ Sacral dimple (Abnormal sacral segmentation)	HP:0008468 (prominent protruding coccyx, HP:0008472)	+	
	Joint Hypermobility	HP:0001382	+ (metacarpophalangeal joint hyperextensibility, HP:0006099) & (hyperextensibility at wrists, HP: 0005072)	+ (fingers & hyperextensibility at wrists (HP:0005072))
	Hip Dysplasia	HP:0001385	-	+
	Spasticity	HP:0001257	+ (lower extremity (HP:0002061)	
	Pectus Excavatum	HP:0000767		
	Kyphosis	HP:0002808	+	+ (thoracic, HP:0002942)
	Scoliosis	HP:0002650	-	+
	Short Neck	HP:0000470	+	
Respiratory			-	+ (H/O asthma (HP:0002099)
Cardiovascular	Structural Defects at Birth		-	- (H/O murmur (HP:0030148) without cardiovascular abnormalities)
Psychiatric	Autistic Behaviors	HP:0000729	+	
	Attention Deficit Hyperactivity Disorder	HP:0007018	+	
	Anxiety	HP:0000739	+	
Other	Intellectual Disability	HP:0001249	+	
	Other Features		-	overlapped toes, genu balgum (HP:0002816)

“+”: presence of a phenotype; “-”: feature absent; NK: not known.

The parents of the two affected siblings are non-consanguineous and are both healthy. The mother was evaluated for phenylketonuria and had normal plasma amino acid levels. The family history does not reveal any members, living or deceased, with phenotypic or syndromic characteristics that resemble the described syndrome, and there is a male cousin who is unaffected. An X-chromosome skewing assay (O'Rawe et al., 2015) revealed that the mother of the two affected boys has skewed, 99:1, X-chromosome inactivation. The grandmother, as well as the aunt of the affected boys, do not show any appreciable X-chromosome skewing, suggesting a newly arising deleterious X-chromosome variant.

Both pregnancies with these male fetuses were complicated by placenta deterioration, and both affected siblings were diagnosed with intra-uterine growth retardation (IUGR) and were delivered through Caesarean section (C-section). The mother denied any alcohol or drug use, nor any exposure to environmental toxins during the course of both pregnancies. The elder boy was born in the 40th gestational week with a birth weight of 2.21 kg and a notable birth defect of aplasia cutis congenita, which was surgically corrected at the age of 4 days old. The younger boy was born in the 37th gestational week with a birth weight of 1.76 kg. A heart murmur was noticed at his birth, but echocardiography confirmed the absence of any further or more serious cardiovascular abnormalities. He was treated with light for neonatal jaundice, and required a feeding tube during the first few days of his life due to difficulties swallowing and digesting food. During the most recent examinations, the younger boy (aged $10^{11/12}$ years) had a height of 129.7 cm (2% tile), a weight of 30.8 kg (19% tile, BMI 18.3 kg/m2), and his occipital frontal circumference (OFC) was 51 cm (4.5th percentile); while his elder brother (aged $11^{11/12}$ years) had a height of 136.8 cm (5% tile), a weight of 26.3 kg (0% tile, BMI 14.1 kg/m2), and his OFC was 49.5 cm (0.2^{th} percentile) at the time.

Brain MRIs of the two brothers demonstrated a remarkably similar constellation of abnormalities (Fig 2.2). In both subjects, there was hypoplasia of the isthmus and splenium of the corpus callosum with thickness falling below the third percentile reported for individuals of the same age. As is often the case with callosal hypoplasia, there was associated dysmorphic configuration of the lateral ventricles and mild lateral ventriculomegaly without positive findings of abnormal CSF dynamics (i.e. no imaging evidence of hydrocephalus). There was also deficiency of the septum pellucidum in both brothers, with the older brother having absence of the posterior two-thirds of the septum pellucidum and the younger brother having complete absence of the septal leaflets. Findings associated with septooptic dysplasia included underdeveloped pituitary glands for their age, deficiency of the anterior falx with mild hemispheric interdigitiation, and question of small olfactory bulbs despite fully formed olfactory sulci. However, the optic nerves appeared grossly normal in size. Finally, there was subjective vermian hypoplasia with the inferior vermis resting at the level of the pontomedullary junction rather than a more typical lower half of the medulla. Pertinent negatives included absence of a malformation of cortical development, evidence of prior injury, or conventional imaging evidence of a metabolic/neurodegenerative process.

Other clinical diagnostic testing performed on both affected siblings in the clinic (see Clinical Methods section) did not reveal any known disorders. Although chromosomal analysis revealed that both boys have the karyotype of 46,X,inv(Y)(p11.2q11.2), this is known to be a normal population variant.

Locating the potential disease-contributory variant

The pool of variants were filtered by Jason O'Rawe based on three disease models: X-linked, autosomal recessive, and autosomal dominant inheritance according to the phenotype segregation pattern seen among the family members. The highest ranking SNVs post variant prioritization included *ZNF41*, *ASB12*, *PION*, and *TAF1* (detailed filtering process please refer to the original paper (O'Rawe et al., 2015)). Jason's analysis of genotyping arrays and the WGS data also led to the identification of a number of CNVs; however, none of them are known to be associated with any phenotype. Hence, we decided to perform Sanger sequencing experiment to validate the allele information of the above four SNVs in all ten family members (Fig 2.4).

Interestingly, it turned out the variants in *ASB12* and *ZNF41* shared by the mother and her two affected sons were also carried by the probands' unaffected male cousin III-3, hence they are unlikely to be the key disease-contributory variants. Moreover, the heterozygous variant in *PION* shared by both of the probands was indeed passed down from their mother who inherited from her mother. However, the probands' grandma is heterozygous for this variant while their mother carried a homozygous allele instead. This indicated there was a de novo mutation event that arose in the mother but is not associated with the patients' symptoms. To confirm this finding, we repeated this validation experiment two times. Based partly on findings in chapter 1, we suspect that this observation is likely due to varying false positive and false negative rates across sequencing and informatics platforms. This validation process demonstrated the benefit of incorporating genetic data from a multi-generational family for improving the accuracy in variant discovery.

This TAF1 variant (p.Ile1337Thr (NP_001273003.1)), which was de novo on the X-chromosome of the mother (II-2) and transmitted to both of her sons, was robustly identified between two different prioritization analysis schemes applied to the WGS data. According to Jason's analysis, it was ranked highest among those tested with VAAST using an X-linked model, and it was ranked by CADD as being within the top 1% most deleterious variants in the human genome with a score of 27. In addition, it was categorized by PolyPhen-218 as being "Probably Damaging" with a score of 0.996, by SIFT19 as "Damaging" with a score of 0.003, by GERP++ 20 with a score of 5.6, by phyloP21 as "Deleterious" with a score of 7.695.

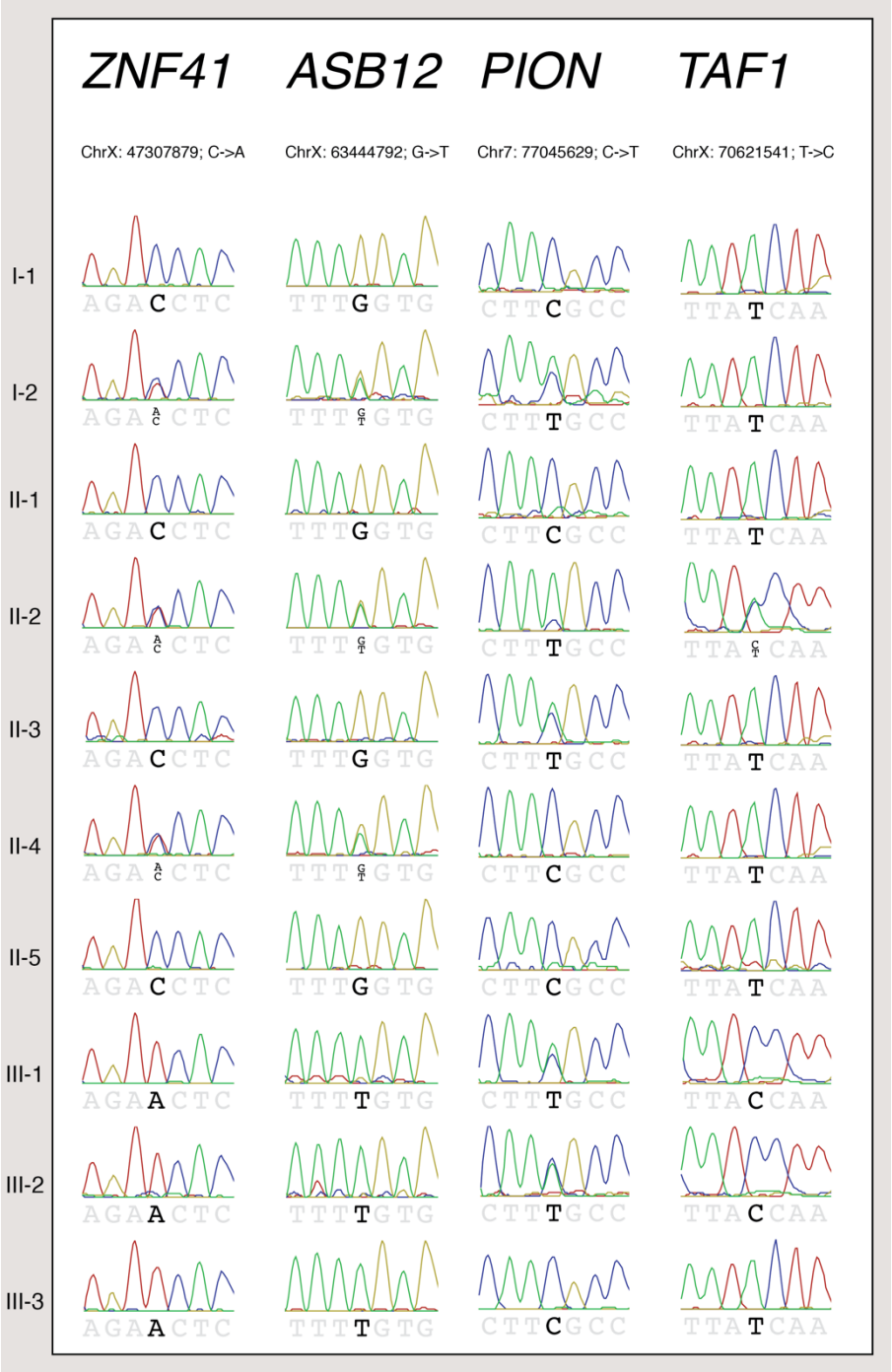


Figure 2.4 Sanger sequencing analysis of 10 family members in project 1 for variants found in *ZNF41*, *ASB12*, *PION* and *TAF1*. Please refer to Figure 2.1 for pedigree information.

TAF1 (TATA-box-binding protein associated factor 1) is part of the TFIID multi-protein complex, which consists of the TATA binding protein (TBP) and 12 additional TBP associated factors (TAFs), and TAF1 is the largest known TFIID associated TAF. It has been shown that TAF1 transcript depletion can lead to aberrant expression affecting hundreds of D. melanogaster genes (Pennington et al., 2013), and studies in rat and mice revealed the importance and relevance of TAF1 expression patterns specific to neuronal tissues (Jambaldorj et al., 2012; Sako et al., 2011). Furthermore, mutations in other TAFs and the TBP have been linked to human neurodegenerative disorders such as X-linked dystonia-parkinsonism (Herzfeld et al., 2013; Makino et al., 2007) as well as in intellectual disability and developmental delay, respectively (Abu-Amero et al., 2010; Hellman-Aharony et al., 2013; Rooms et al., 2006). However, the only link of TAF1 in human disease pathogenesis was discussed in a recent report where TAF1 was nominated as a candidate gene for intellectual disability (Hu et al., 2015). The protein change in TAF1 in our family occurs within a linker region N-terminal to two bromo-domains and C-terminal of a co-factor (TAF7) interacting domain (O'Rawe et al., 2015). This linker region is highly conserved in multicellular eukaryotes, suggesting an important functional role for this motif.

Finding more patients with TAF1 pathogenic variants

After pinpointing the potential disease-contributory TAF1 variant in our probands, we published our initial findings on the bioRxiv preprint server (http://biorxiv.org/), with the hope to find other similar patients. We described in detail about the clinical features of this new syndrome to help physicians and researchers identify patients with similar phenotype (O'Rawe et al., 2015). Luckily, this approach worked, and it eventually helped us gather a study cohort of 14 affected individuals from 11 unrelated families worldwide. 12 of these individuals (from nine

unrelated families) contain single-nucleotide changes in *TAF1*, and two (from two unrelated families) have large duplications involving *TAF1*. Strategies used for identifying candidate disease-related sequence variation among these groups included whole genome sequencing, whole exome sequencing, targeted gene panel sequencing, and microarray-based strategies.

Commonly shared features of these individuals include global developmental delay, intellectual disability, characteristic facial dysmorphologies, generalized hypotonia, joint hypermobility, hearing impairment, microcephaly, and hypoplasia of the corpus callosum. Shared facial dysmorphologies include prominent supraorbital ridges, down-slanted palpebral fissures, sagging cheeks, a long philtrum, low-set and protruding ears, a long face, a high palate, a pointed chin, and anteverted nares. Interestingly, the characteristic gluteal crease with a sacral caudal remnant was widely shared among these probands. Some of these individuals have progressive symptoms, and one died of severe cardiopulmonary insufficiency attributed to an infection. Of note, probands with duplications demonstrate severe and progressive neurological impairment, but do not share some of the more common features with the rest of the probands (O'Rawe et al., 2015).

This collaborative effort from over 50 scholars around the world of identifying a new syndrome using NGS illustrated the power of sharing both phenotype and genotype information in expediting the discovery of the genetic basis of human disease.

Project 2. A complex pedigree with dysautonomia-like symptoms

Clinical presentation and family history

This Utah pedigree K10031 consists of 14 individuals from three generations (Fig 2.5) with various medical conditions including Prader-Willi syndrome (PWS), hereditary hemochromatosis (HH), dysautonomia-like symptoms, Tourette syndrome (TS) and other illnesses. The two probands we discuss in detail below come from two nuclear families in this extended pedigree. All clinical information was collected by various physicians, but was systematically analyzed and organized by myself.

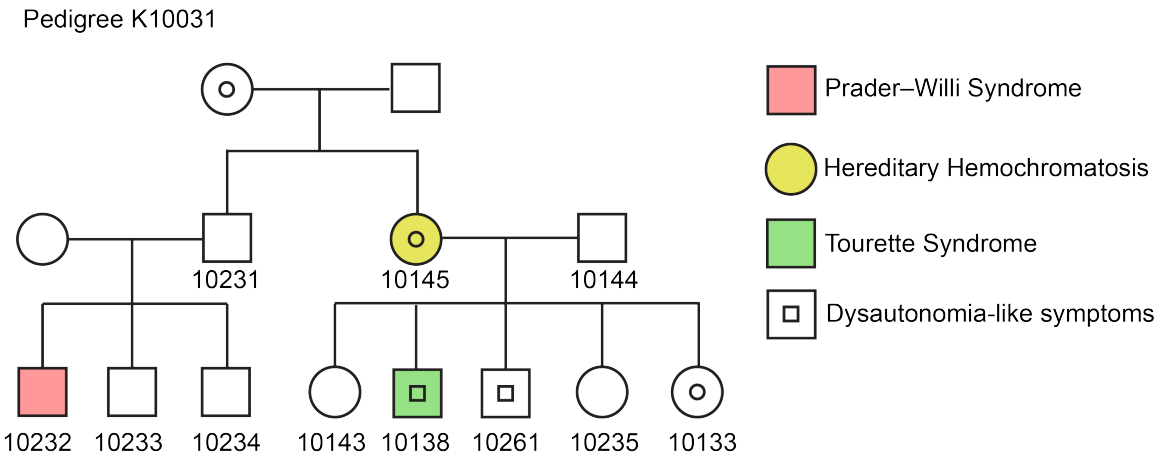


Figure 2.5 Pedigree of the studied family in project 2. All samples from individuals with a number underneath underwent WGS except for K10031-10234 and K10031-10261.

Proband K10031-10232

Proband K10031-10232 is a 25-year-old (25 y.o.) male. He is the son of a Caucasian farther (K10031-10231), and an Asian mother (who did not participate in the study). He has two older male siblings, namely K10031-10233 and K10031-10234. This proband was diagnosed with PWS at 11 months old, and has dysmorphic facial features including a narrow forehead, downslanted palpebral fissures and almond-shaped eyes. Since the PWS diagnosis, his behavior has been assessed in great detail, and the following diagnoses have been given: obsessive-compulsive disorder (OCD), depression, anxiety disorder, pervasive developmental disorder (PDD), hyperphagia, trichotillomania, and daytime hypersomnolence. He has an IQ ranging between 60-65, categorized as mild mental retardation. He also has diagnoses of mild dysarthria, obstructive sleep apnea syndrome (OSAS), and severe scoliosis. The latter has been corrected surgically. He has also undergone orchiopexy, tonsillectomy, and adenoidectomy. His physical exam is otherwise unremarkable. He has denied having significant psychotic symptoms, including auditory or visual hallucinations, delusions, ideas of grandiosity, or paranoid ideation. See Table 2.2 for a list of clinical phenotype features collected from this proband.

Table 2.2 Clinical presentation of proband K10031-10232

Prenatal History	HPO #
Cesarean section	0011410
Gestational diabetes	0009800
Oligohydramnios	0001562
Premature birth (36 weeks gestation)	0001622
Development and Growth	
Delayed speech and language development	0000750
Dysarthria	0001260
Growth hormone deficiency	0000824
Poor fine motor coordination	0007010
Mild intellectual disability	0001256
Facial Features	
Abnormal facial shape	0001999
Almond-shaped eyes	0007874
Downslanted palpebral fissures	0000494
Narrow forehead	0000341
Other Physical Features	
Cryptorchidism	0000028
Excessive daytime sleepiness	0002189
Infantile muscular hypotonia	0008947
Obstructive sleep apnea syndrome	0002870
Scoliosis	0002650
Behavior Features	
Aggressive behavior	0000718
Anxiety	0000739
Depression	0000716
Flipping pages in books	NL
Hair-pulling	0012167
Hyperesthesia	0100963
Impaired ability to form peer relationships	0000728
Impaired social reciprocity	0012760
Inflexible adherence to routines or rituals	0000732
Irritability	0000737
Lack of insight	0000757
Lack of peer relationships	0002332
Low frustration tolerance	0000744
Nail-biting	0012170
Obsessive-compulsive disorder	0000722
Pain insensitivity	0007021
Polyphagia	0002591
Poor eye contact	0000817
Restrictive behavior	0000723
Short attention span	0000736
Skin-picking	0012166

BMI, body mass index; NL, not listed.

The genomic analysis of proband K10031-10232 performed by Han Fang confirmed de novo type I PWS deletions (Cassidy et al., 2012; Christian et al., 1995) in the chromosome regions from 15q11.2-15q13.1, which his father and his brothers (K10031-10233 and K10031-10234) do not carry. This indicates that the proband only carries the maternal alleles in those regions (proband's mother refused to participate, hence no sequencing data was available to confirm). The Phenomizer analysis of this proband also ranked the PWS diagnosis as the highest priority and revealed how genes in the deletion regions are linked with the phenotypes represented by HPO terms. This result further supported the fact that highly specific and annotated phenotype information can yield accurate diagnoses. Since PWS is a genetic syndrome with known genetic etiology, it served as a positive control among a range of diseases of unknown (or controversial) genetic architecture in this study.

Proband K10031-10133

Proband K10031-10133 is a 26 y.o. female, born to a Caucasian mother (K10031-10145) and a Caucasian father (K10031-10144). She is the eldest child amongst her two sisters and two brothers. Prior to age 18, K10031-10133 had a fairly unremarkable medical history. Arthralgia and episodes of fatigue and dizziness started at around 18 years of age. At age 20, she started to have refractory syncopal events, which led to multiple body injuries. During the same period of time, she also developed postural orthostatic tachycardia syndrome (POTS), heart palpitations, gastroparesis, urinary incontinence, diplopia, and seizures. In addition, she reported experiencing auditory and visual hallucinations. She underwent dysautonomia evaluation, which included a positive tilt table test. Other tests revealed unusual changes to her optic disks but without an elevated intraocular pressure, and nonspecific findings on her brain MRI, including a subtle focus of T2 signal abnormality involving the subcortical white matter of the right parietal lobe

without associated enhancement. See Table 2.3 for proband K10031-10133's clinical phenotype list with HPO annotations.

Table 2.3 Clinical presentation of proband K10031-10133's family

Pedigree K10031		10133	10145	10235	10261	10138	10143
General Information							
Age (years)		26	45	24	22	20	18
Gender		F	F	F	M	M	F
Clinical Manifestations	HPO #						
Developmental/Growth							
Dyslexia	0010522	-	-	-	+	-	-
Cardiovascular							
Bradycardia	0001662	+	-	-	-	+	-
PFO	0001655	+	-	-	-	-	-
Syncope	0001279	+	+	-	+	+	-
Tachycardia	0001649	+	+	-	-	-	-
Eyes							
Astigmatism	0000483	+	-	-	-	-	-
Diplopia	0000651	+	-	-	-	-	-
Myopia	0000545	+	-	-	-	-	-
Optic disks changes	NF	+	-	-	-	-	-
Peripheral vision	NF	+	-	-	-	-	-
Prominence to the eyes	NF	+	+	+	+	+	+
Gastrointestinal							
Acid reflux	0002020	-	-	+	-	-	-
Anorexia	0002039	-	+	-	-	-	-
Gastroparesis	0002578	+	-	-	-	-	-
Nausea	0002018	+	-	-	-	-	-
Gynecologic & Genitourinary							
Irregular periods	NF	+	-	-	N/A	N/A	-
Urinary retention	0000016	+	-	-	-	-	-
Urinary incontinence	0000020	+	-	-	-	-	-
Hematologic/Lymphatic/Immunologic							
Adenopathy	Lymphoadenopathy*: 0002716	+	-	-	-	-	-
Musculoskeletal							
Arthralgia	0002829	+	-	-	-	-	-
Arthritis	0001369	-	-	-	-	-	+
Cyanotic lower extremities	0001063	+	-	-	-	-	-
Joint Stiffness	0001387	+	-	-	-	-	-
Muscle weakness	0001324	+	-	-	-	-	-
Osteoporosis	0000939	-	+	-	-	-	-
Neurological							
Apraxia	0002186	+	NK	NK	NK	NK	NK
Arthritis	0001369	+	-	-	+	-	+
Auditory hallucinations	0008765	+	-	-	-	-	-
Concussion	NF	+	-	-	-	-	-
Convulsions	NF	+	-	-	-	-	-
Dizziness	NF	+	+	-	+	-	+
Dysarthria	0001260	+	-	-	-	-	-
Fatigue	0012378	+	+	-	-	-	-
Frequent falls	0002359	+	-	-	-	+	-
Headache	0002315	+	NK	NK	NK	NK	NK
Heat intolerance	0002046	+	+	-	-	-	-
Ischemic stroke	0002140	+	-	-	-	-	-
Migraine	0002076	+	-	+	+	+	+

Numbness	Paresthesia*: 0003401	+	-	-	-	-	-
Seizure	0001250	+	-	-	+	-	-
Tremor	0001337	+	+	+	+	+	+
Visual hallucinations	0002367	+	-	-	-	-	-
Respiratory							
Asthma	0002099	+	+	+	+	+	+
Dyspnea	0002094	+	NK	NK	NK	NK	NK
Psychiatric							
ADHD	0007018	-	-	-	+	+	-
Anxiety	0000739	+	+	+	-	+	+
Depression	0000716	+	+	-	-	-	-
Dissociative identity disorder	NF	-	+	-	-	-	-
Obsessive-compulsive behavior	0000722	-	+	-	-	+	-
Tourette syndrome	ORPHANET:856	-	-	-	-	+	-

"+": presence of a phenotype; "-": feature absent; N/A, not applicable; NF, not found; NK, not known. Terms with * refer to the synonym of the disease or feature in question as categorized in the HPO Phenomizer tool.

As for her family history, there are some noticeable symptoms that are shared by all her siblings and her mother, including dysautonomia-like symptoms such as dizziness and fainting, as well as tremors and asthma. In addition, anxiety, attention deficit, arthritis, dyslexia, gastroesophageal reflux, seizures and Tourette's syndrome (TS) are other diagnoses found among her siblings. Her mother (K10031-10145), on the other hand, has hereditary hemochromatosis (HH) and obsessive-compulsive disorder (OCD) traits. Her father has significant migraines, gastroesophageal reflux, hiatal hernia, and right sensorineural hearing loss.

Analysis of dysautonomia-like symptoms

Analysis of WGS data showed that none of the family members with dysautonomia-like symptoms carry any previously reported variants in *IKBKAP* that are implicated in the autosomal recessive transmission of FD (Dietrich & Dragatsis, 2016; Norcliffe-Kaufmann et al., 2016), which is also called hereditary sensory and autonomic neuropathy type III (HSAN- III). The WGS data have effective sequence coverage (> average coverage 40×) for this gene, but no novel rare variants were identified. Notably, both the mother (K10031-10145) and the male proband (K10031-10138) carry heterozygous variants of p.H604Y and p.G613V in the protein product of *NTRK1*, which has been proven to contribute to HSAN-IV (congenital insensitivity to pain with anhidrosis). HSAN-IV is a disease closely resembling FD (HSAN III), and is characterized by a lack of pain sensation, anhidrosis, unexplained fever since childhood, and self-mutilating behavior (Indo et al., 1996; Swanson, 1963). Both variants are located within the intracellular tyrosine kinase domain of the encoded protein, but neither sites are conserved. Both variants have also been reported before in healthy individuals, so they are considered to be polymorphisms in the population and seem to be in linkage disequilibrium (Cargill et al., 1999; Gimm et al., 1999; Greenman et al., 2007; Indo, 2001; Shatzky et al., 2000). The mother's

brother (K10031-10231, unaffected) also carries these two variants, so this provides further evidence that they are likely to be polymorphisms. Most importantly, neither variant is present in the proband K10031-10133, who reported the most severe dysautonomia-like symptoms.

In addition to the *NTRK1* variant, manual filtering (performed by Han Fang) found seven other putative variants in *PLCG2*, *ATXN2*, *VWA8*, *LRRIQ1*, *MYO1H*, *OR1J4*, and *RFX4* which follow a dominant inheritance model. Variants in *PLCG2*, *ATXN2*, and *VWA8* were previously reported to be associated with certain disease phenotypes, including cold-induced urticaria, antibody deficiency, susceptibility to infection and autoimmunity, spinocerebellar ataxia type 2, celiac disease, and susceptibility to amyotrophic lateral sclerosis (Hunt et al., 2008; International Multiple Sclerosis Genetics Consortium et al., 2011; Ombrello et al., 2012). However, the variants we identified in this family are not the same variants in the literature, and all of these predicted diseases have only partially overlapping manifestations with dysautonomia-like symptoms. For the rest of the four genes mentioned above, *LRRIQ1*, *MYO1H*, *OR1J4*, and *RFX4*, there has been, to our knowledge, no reports published to date discussing any variants in these genes associated with human disease. Therefore, the functional impact of these variants remains unclear.

HFE p.C282Y variant and hereditary hemochromatosis

Han's analysis revealed the mother (K10031-10145) with HH is homozygous for the p.C282Y variant in HFE, which is consistent with her molecular genetic assay results. However interestingly, her brother (K10031-10231) who has a normal clinical test also carries the same variant, even though male p.C282Y homozygotes are considered more likely to develop iron-overload–related diseases due to the lack of the iron clearance events like menstruation and

pregnancy in women (Allen et al., 2008) . This is in line with the fact that even family members can have variable expressivity of disease, including different onset ages, etc. This instance with the brother and the sister again highlights the point that the phenotypic expression of a given mutation in *HFE* may vary widely, influenced in part by unidentified modifier loci (Hanson et al., 2001; McLaren et al., 2015; Pelucchi et al., 2012; Pietrangelo, 2004; Stickel et al., 2014; Triess et al., 2012; Valenti et al., 2012).

2.4 Discussion

In our second study, we tried to pinpoint the genetic basis of patients' dysautonomia-like symptoms without much success, regardless of our extensive efforts of combining the use of multiple analysis tools to increase our chances. In fact, the overall success rate of using clinical genomics for the identification of a causative variant is still low, approximately 25% (Yang et al., 2013). The reasons are many. First, the lack of precise and specific phenotype data, with us being unable to access patients' medical records, proved to be a great challenge during our study design and variant prioritization process. Besides, if we can obtain the genome information of the mother's mother (who also had dysautonomia-like symptoms), our chance of finding the disease-contributory variant will also be enhanced, since a filtering strategy based on three generations will further help eliminate false positive and negative variants. Furthermore, the lack of clinical actionable variant database is another hurdle of accurately estimate and interpret the pathogenicity of variants (Yang et al., 2013). Databases such as ClinVar (Landrum et al., 2014) had been developed to meet such purpose; however, studies (Amendola et al., 2016; Rehm et al., 2015) have shown that 17-42% of variants in ClinVar can be described as having a different potential for pathogenicity by different submitters, the category can vary from "pathogenic or likely pathogenic," "uncertain significance," to "likely benign or benign", which can be really

confusing and worrying especially for clinical decision-making. Hopefully, the level of consensus for pathogenicity will be improved in the near future with more and more evidences accumulated from biological functional studies.

Meanwhile, our results also suggest the genetic inheritance of dysautonomia-like symptoms in this pedigree may not consist of only one high-effect size mutation, but rather could be polygenic and/or environmentally influenced. It is possible that multiple variants including those we mentioned above are acting together or in conjunction with modifiers in these individuals' genomes to give rise to a spectrum of complex clinical manifestations. Of course, we cannot exclude the possibility that we might have missed some variants, including possibly non-coding variants, and we expect that the future phenotyping, sequencing, and collation of data from millions of people will reveal associations that are not currently known.

On the other hand, we also illustrate the necessity of taking caution when interpreting a positive clinical genomics finding. The first example is the p.C282Y variant found in the clinically unaffected brother in our second study. We still need to realize the limitation of our understanding of the pathogenesis and molecular basis of many diseases. In contrast to studies that have searched for the "causal" gene, some have reported that genetic variations can instead have large effects on phenotypic variability, suggesting underlying genomic complexity from multiple interacting loci (Corbett-Detig et al., 2013; Mackay et al., 2009; Massouras et al., 2012; Zuk et al., 2012). Understanding such diseases thus requires probabilistic thinking about the risk of developing the clinical manifestation, rather than deterministic genotype-phenotype "causation" (Freund et al., 2013; Moczulski et al., 2001; Thornton-Wells et al., 2004). Another example is during the analysis of identifying variants related to dysautonomia-like symptoms, we actually found an in-frame insertion in *POU4F1* that was present in all affected people in the

family (Fig 2.6). However, follow-up Sanger sequencing result showed that other unaffected people in the family also carry this variant, which was missed by WGS due to low coverage in these controls. This is an example of the mis-annotation of the reference genome (since this "insertion" sequence is not seen in the reference genome), which situation has been reported by other groups as well (Rosenfeld et al., 2012). Our study highlights the importance of genotype refinement and use of integrative pipelines, especially considering the incompleteness and inadequate representation of human diversity in the current reference sequence (Kidd et al., 2010; Li et al., 2010).

Gene	Chr	Position	Region	Transcript Variant	Protein Variant	Inferred Activity
POU4F1	13	79176323	Exonic	c.484_486dipGGC	p.162_163_insG	Loss

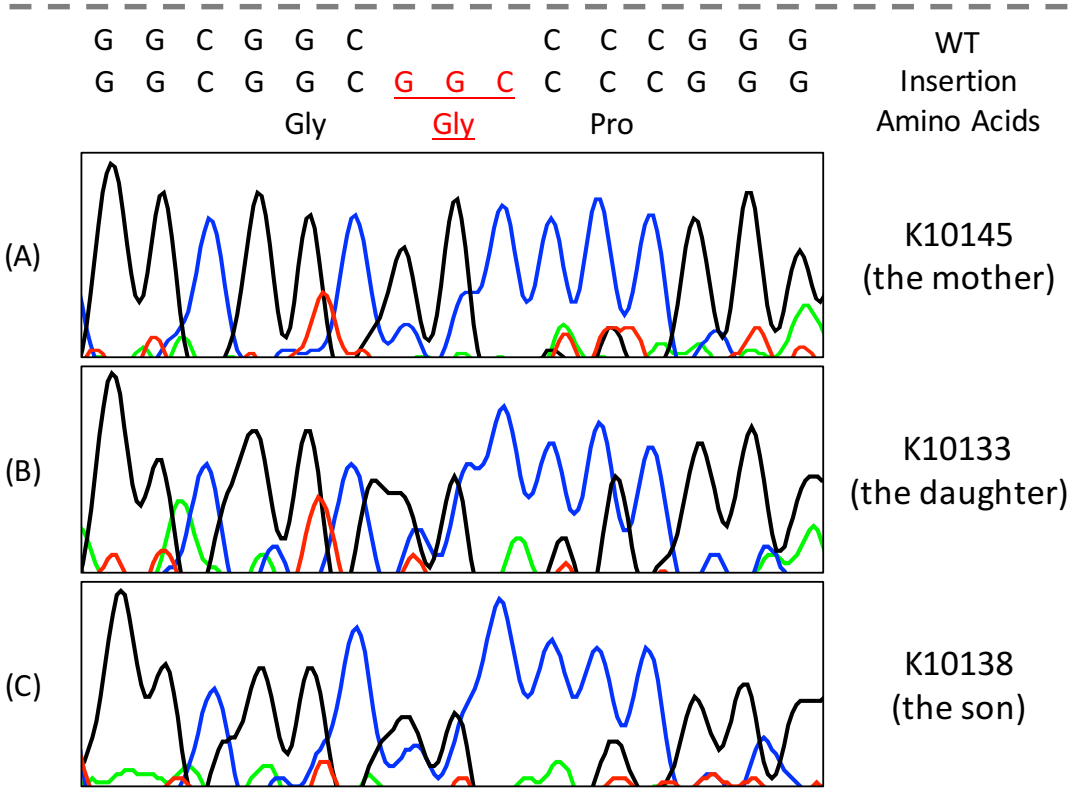


Figure 2.6 Sanger sequencing results of the variant found in *POU4F1*. Results confirmed the homozygous insertion in exon 2 of *POU4F1* in (A) the mother K10145, (B) the oldest daughter (proband) K10133, and (C) the son K10138.

As we know, the evidence obtained from a functional study of the disease-assoicated variant is the key to demonstrate the causality and understand the mechanism of the disease pathogenesis. In our TAF1 story, collaborative work with taf1 knockdown in developing zebrafish embryos using morpholino and CRISPR/Cas9 revealed a possible disease-related phenotype, a 10% reduction of the area of the optic tectum (i.e. microcephaly), although a rescue experiment failed to show any differences between WT and mutated TAF1 constructs (O'Rawe et al., 2015). Nonetheless, this observation provides additional evidence supporting the important role of TAF1 in the development of central nervous system. Such functional studies can be carried out using a variety of models, including patients' cells and human induced pluripotent stem cells (hiPSCs), which I will illustrate in detail in the following two chapters in the context of a rare genetic disorder called Ogden syndrome.

2.5 Concluding Remarks

In conclusion, I illustrated here, with the TAF1 story, the opportunity to use clinical genomics as a tool to accelerate the discovery of the genetic basis of diseases. And in the meantime, our second project demonstrates that NGS might not always successfully pinpoint the exact disease relevant variants due to disease complexity and sometimes the limitation of the cohort size and clinical information.

In both of our projects, we carried out extensive analysis of patients' clinical data, and used standardized vocabulary (HPO) to describe, compare and share phenotypes. As the field of precision medicine advances, such an approach will enable a more efficient capture of phenotypic information and allow for the use of computational algorithms to search for phenotypic similarity between genomics studies. Furthermore, we also highlight the integrative use of WGS and HPO data to help eliminate false positives and maximize their potential for clinical implementation. Lastly, we demonstrate the benefits of studying large related cohorts, where intra-familial relationships enable more rigorous variant filtering and identification of true positive alleles that might be contributing to a disease phenotype.

Lastly, we also recommended to interpret and convey a positive clinical genomics finding with caution (Boyd et al., 2014; Richards et al., 2015), due to our limited knowledge of disease pathogenesis and the underrepresentation of the current reference genome for human variation.

Chapter 3

Ogden syndrome is linked to downstream N-terminal acetylation defects

3.1 Introduction

Ogden syndrome

Ogden syndrome was reported in 2011 by a team of researchers from Utah led by Dr. Lyon (Rope et al., 2011). Through X chromosome exon capture and sequencing in combination with a variant annotation, analysis, and selection tool called VAAST, the team successfully

pinpointed this disease-contributory variant in NAA10 c.109T>C. The same variant was later found and confirmed in a second unrelated family with probands presenting a similar phenotype as patients in the first family. Ogden syndrome is an extremely rare, infantile lethal X-linked disorder. To date, there are only eight Caucasian males (with this exact mutation) from the above two families of European descents identified worldwide (Fig 3.1) (although other NAA10 mutations have been identified (see discussion)). Patients with Ogden syndrome are characterized by a distinct combination of dysmorphic features, cardiac dysfunction, hypotonia, global developmental delays and growth failure. The average life span of eight affected male probands was 10.2 months (ranging from 4.5-16 months). Death was often sudden and associated with cardiac arrhythmias. Female carriers of Ogden syndrome do not have any confirmed clinical findings.

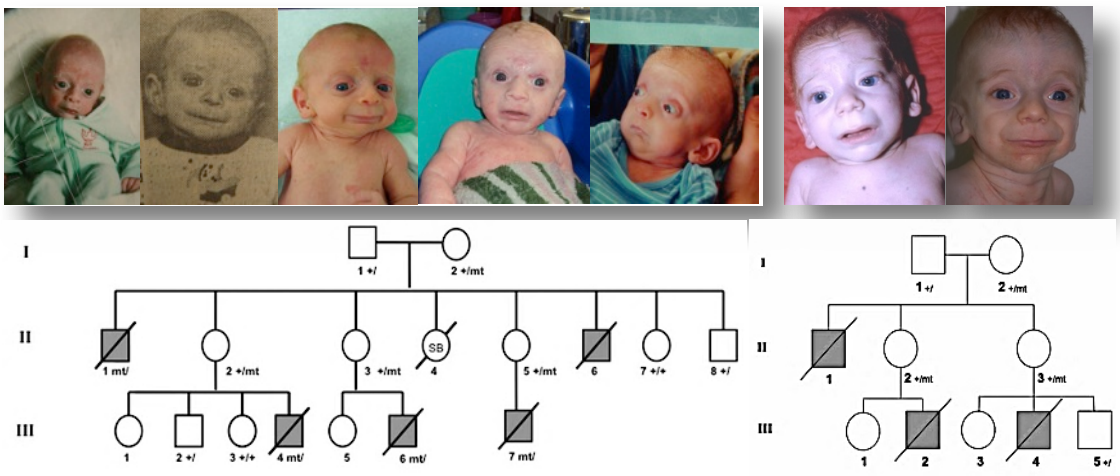


Figure 3.1 Ogden syndrome pedigrees from two unrelated families with eight probands. The pedigree on the left is the Utah family, which was studied in depth in this dissertation. Pedigrees and some of the images were re-used from the original paper with permission (Rope et al., 2011).

NAA10, NatA and protein N-terminal acetylation

The gene which harbors the variant for Ogden syndrome is NAA10 on chromosome X, which encodes N-α-acetyltransferase 10 (NAA10). NAA10 is the catalytic subunit of human N-terminal acetyltransferase A (NatA), one of the seven NATs (Nat A-G) that have been identified in eukaryotes to date (Aksnes et al., 2016). The function of NATs is to transfer an acetyl moiety from acetyl coenzyme A (Ac-CoA) to the primary α-amino group of a nascent polypeptide co-translationally and post-translationally. Proteomics studies showed that 80-90% of human proteins are Nt-acetylated (Arnesen et al., 2009); however, the research on biological implications of Nt-acetylation (NTA) in humans is still in its infancy. Current knowledge of the functional role of protein NTA was mostly gained in yeast models, which includes regulating protein turnover through N-end rule pathway (Hwang et al., 2010; Shemorry et al., 2013), mediating protein interactions and complex formation (Arnaudo et al., 2013; Monda et al., 2013), determining protein subcellular localization (Behnia et al., 2004; Dikiy & Eliezer, 2014), affecting protein folding and aggregation (Holmes et al., 2014; Raychaudhuri et al., 2008), as well as some most recent evidence supporting a potential interplay with other protein modification such as ubiquitination and methylation (Scaglione et al., 2013; Schiza et al., 2013).

The NatA complex consists of NAA10 (Ard1, catalytic subunit) and NAA15 (Nat1, the auxiliary subunit), together with the attachment of Naa50 and HYPK (Huntingtin-interacting protein K) (Arnesen et al., 2010; Gautschi et al., 2003; B. C. Williams et al., 2003). Each NAT has its unique substrate specificity, and NatA favors protein N-terminus (the second amino acid residue) starting with A-, S-, T-, V-, C-, and sometimes G, with their initiator methionine (iMet) first being removed by methionine amino peptidases (MetAPs) (Arnesen et al., 2009; Mullen et

al., 1989). It is estimated that 38% of entire proteome is Nt-acetylated by NatA, whereas NatB/C/E/F together acetylate another 42% of the Nt-acetylome, making NatA the major NAT in humans (Aksnes et al., 2016). NatA has been shown to be associated with ribosomes to perform co-translational protein Nt-acetylation (Gautschi et al., 2003; Polevoda et al., 2008). However, there is also evidence showing that NAA10 can display NAT activity independently of the NatA complex. But when functioning on its own, it recognizes a different spectrum of substrates (Aksnes et al., 2015; Evjenth et al., 2009; Liszczak et al., 2013; Petra Van Damme et al., 2011).

Ogden syndrome is the first known genetic disease directly associated with protein NTA pathway, thus it provides a unique model to expand our knowledge on NTA and its role in human development, physiology, and pathophysiology. In the original report of Ogden syndrome, it was shown that NAA10 S37P mutant displays a reduction of NAT activity *in vitro* to various degrees (20-80%) compared to the WT (Rope et al., 2011). Hence, the authors hypothesized that the severe clinical phenotype of Ogden syndrome patients is due to the reduction of *in vivo* NTA of a subset of NatA substrates, which play crucial roles in development and physiological processes.

The work shown in this chapter aims to fully analyze the phenotypic data of Ogden syndrome patients, and investigate the functional impact of the hemizygous NAA10 S37P variant in patient's cells to evaluate the hypothesis above.

3.2 Materials and Methods

Ethics and consent

The sample collection (performed by other people) for fibroblasts and cell line creation was under protocols approved by the Institutional Review Board at the University of Utah and Children's Hospital of Philadelphia (CHOP) and Cold Spring Harbor Laboratory (CSHL). De-identified aliquots of cell lines were distributed for use in the experiments described. Written informed consent was obtained from individuals and families of study for sample collections, publication and use of facial photography. All experiments were performed in accordance with the approved guidelines and regulations.

Cell culture

Primary WT control fibroblasts include BJ fibroblasts (ATCC, # CRL-2522, male newborn's foreskin origin, WT-0), HDFn fibroblasts (Thermo Fisher Scientific, # C-004-5C, male newborn's foreskin origin, WT-1, 2, 3), and an in-house sample (6 year-old male skin-punch biopsy origin, WT-5). Primary Ogden syndrome patient's fibroblasts are from individual III-6 of the Utah family, isolated from a skin-punch biopsy at the University of Utah. Primary fibroblasts were maintained in Eagle's minimal essential medium (EMEM, Lonza) supplemented with 10% fetal bovine serum (FBS, HyClone) and 2 mM L-glutamine (Thermo Fisher Scientific, # 25030081), and cultured at 37 °C under 5% CO2. Human embryonic kidney 293 cells (HEK293, CSHL in-house sample) were cultured in Dulbecco's Modified Eagle's Medium (DMEM, Lonza) supplemented with 10% FBS (Thermo Fisher Scientific, # SH30070.03) and 2 mM L-glutamine (Thermo Fisher Scientific, # 25030081), and cultured at 37 °C under 5% CO2.

Cell line authentication

The in-house cell line used in this study (HEK293) was authenticated through ATCC via Short Tandem Repeat (STR) analysis to validate its identification. The in-house sample is an exact match to ATCC cell line CRL-11268 (293T/17). ATCC cell line CRL-11268 was derived from parental cell line CRL-1573 (293).

Antibodies

Primary antibodies include BrdU antibody (EMD Millipore, # QIA58), mouse anti-V5 (Thermo Fisher Scientific, # R960-25), rabbit anti-GAPDH (Santa Cruz, # sc-25778), mouse anti-Myc (Thermo Fisher Scientific, # R950-25), rabbit anti-NAA10 (Protein Tech, # 14803-1-AP), mouse anti-NAA15 (abcam, # ab60065). Secondary antibodies include AlexaFluor488 goat anti-mouse (Thermo Fisher Scientific, # A-11029) and AlexaFluor594 goat anti rabbit (Thermo Fisher Scientific, # A-11037) for immunofluorescences and anti-mouse IRDyse 800CW (LI-COR, # 926-32210) and anti-rabbit IRDye 680RD (LI-COR, # 926-68071) for western blotting.

Growth curve

20K per well of primary fibroblasts WT-0, WT-1, and Ogden S37P were plated in 12-well plates, and left overnight to ensure attachment (2 plates in total with 24 wells of cells were plated on day 0, with three replicated for each day). The cells were harvested with trypsin and counted every day for 8 days. Representative data are shown as the mean ± standard deviation of three replicates. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$ (two-tailed Student's *t*-test).

BrdU cell proliferation

50K of primary fibroblasts WT-1 and Ogden S37P at passage 4 were plated on each cover slip coated with matrigel (Corning, # 354230) and cultured in a 12-well plate. To measure BrdU incorporation, the cells were incubated for 5 hrs with BrdU label (EMD Millipore, # QIA58, 1:2000) at 46hrs after seeding, then fixed with 4% paraformaldehyde for 15 min at room temperature (RT), washed in PBS, treated with 0.07 N NaOH for 15 min at RT to denature the nuclear DNA, and blocked in 10% goat serum, 1% BSA and 0.3% Triton X-100 for 15 min at RT before staining. The cells were then stained with a BrdU antibody (1:100) for 1 h at RT, washed and incubated in secondary antibody (AlexaFluor goat anti-mouse 488, 1:1000) for 1 h at RT, protecting from light. Hoechst 33342 was used as a nuclear stain (Thermo Fisher Scientific, # R37605). Quantitative analysis was performed by counting, for each cell types, the percentage of BrdU-positive nuclei on a total of about 1,000 cells, randomly observed in 15 microscopic fields from WT-1 fibroblasts and 22 microscopic fields from Ogden S37P fibroblasts. All data are shown as the mean ± standard deviation of 1,000 replicates. ***$p < 0.001$ (*Pearson's* chi-squared test). This experiment was repeated four times.

XTT cell viability

Primary fibroblasts were seeded at 5,000, 15,000, 50,000 per well in a 96-well plate and incubated overnight for attachment. At 24 h post-seeding, fresh made XTT detection solution was added to each well of 96-well plates according to the manufacturer's instruction (XTT Cell Viability Kit, Cell Signaling Technology, # 9095). The absorbance at 450 nm was measured with a SpectraMax Plus 384 Microplate Reader (Molecular Devices) at t=0, 1, 2, 3, 4, 5, 17 h after

adding the XTT reagents. The average of quadruplicate measurements is shown with error bars representing the standard deviations. This experiment was repeated four times.

Co-immunoprecipitation and Western blots

Co-IP was performed in collaboration with Max Doerfel. 8×10^5 HEK293 cells were seeded per well in 6-well plates and incubated at 37°C for 24 h. Cells were transiently transfected with polyethylenimine (PEI) with 2 µg pcDNA3.1/V5-His NAA10-WT, pcDNA3.1/V5-His NAA10-S37P, pcDNA3.1/Myc-His NAA15-WT and/or empty control vectors using PEI. Cells were lysed 48 h post-transfection in 200 µl lysis buffer (PBS, 0.2%, v/v, Triton X-100, complete protease inhibitor cocktail (Roche)). Cellular debris was pelleted at 20,800 g for 10 min at 4°C. The protein concentration was determined using advanced protein assay (Cytoskeleton, Inc.) and 600-800 µg total protein was used for IP. Cell lysates were incubated with 1 µg anti-V5 or anti-Myc antibody for 1 h under constant agitation at 4°C. Immune complexes were precipitated with 30 µl 1:1 slurry Protein-A sepharose (Invitrogen) for 30 min at 4°C. Beads were pelleted by centrifugation at 2,700 g for 2 min, washed three times with 300 µl lysis buffer, and protein complexes were eluted with 50 µl 2× SDS-sample buffer and analyzed by SDS-PAGE and Western blotting.

For Western blots, protein samples were run on 10% SDS-PAGE gels and transferred to PVDF membrane (Thermo Scientific, # 88518). The membranes were blocked with Odyssey Blocking Buffer (LI-COR Biosciences, # 927-40000) overnight at 4℃ and then probed with indicated primary antibody at desired dilution in 1:1 Odyssey Blocking Buffer and 0.1% Tween in Tris-buffered saline (TBS-T) for 1 h at room temperature. The membranes were rinsed for 10 minutes 3 times with TBS-T at room temperature. Indicated secondary antibody was applied to

the blot at desired dilution in 1:1 Odyssey Blocking Buffer and TBS-T and incubated for 1 h at room temperature. Western blot image detection and quantification were performed using the LI-COR Odyssey. Protein quantification was performed using the Odyssey Image Studio software (LI-COR Biosciences). This experiment was repeated four times.

Mitochondrial staining

Passage number matched (P7) WT control fibroblasts and S37P patient fibroblasts were seeded on cover slips and incubated overnight to ensure attachment. To label mitochondria, cells were washed with PBS twice and then incubated in 50 nM MitoTracker Red CMXRos (Molecular Probes, # M7512, shared aliquot given by Dr. Daniel Bogenhagen at SBU) at 37°C for 90 min. Then, cells were washed with PBS twice, fixed with 4% PFA, and washed again twice with PBS before being examined under a confocal laser scanning microscope (Zeiss LSM 780). At least 6 fields of view (FOV) of WT1 and 15 FOVs for S37P cells were surveyed and reviewed for comparison in two independent experiments.

Genomic DNA extraction from cell cultures

Adherent cells were trypsinized and collected from the culture dishes by centrifuging for 5 min at 300 g. Cell pellet was resuspended in 200 μl PBS, and then 20 μl proteinase K. was added. Then a standard total DNA extraction protocol was performed according to manufacturer's instructions using DNeasy Blood & Tissue Kit (Qiagen, # 69504). Genomic DNA was quantified by NanoDrop 1000 spectrophotometer and stored at -20℃ until using.

Sanger sequencing

Polymerase chain reaction (PCR) primers for evaluating Ogden syndrome allele (c.109) in NAA10 were designed using Primer3 (http://primer3.sourceforge.net) to produce an 845 bp amplicon. Sequences of the primer pair designed and used are 5'GACTGCGCCTTCACGATCCG and 5'TCCGGGCTGAGTGGCTTGG. Primers were obtained from Sigma-Aldrich, and tested for PCR efficiency with an in-house DNA sample using Phusion Flash II DNA Polymerase (Thermo Fisher Scientific, # F548S). The optimized PCR reaction was then carried out using desired DNA as template. PCR products were visually inspected for amplification efficiency using agarose gel electrophoresis and were purified using the QIAquick PCR Purification Kit (QIAGEN). Purified products were then diluted to 5-10 ng/μL in water for use with a in-house ABI 3700 sequencer. The resulting ∗.ab1 sequence files were loaded into the CodonCode Aligner V5.1.2 for analysis. All sequence traces were manually reviewed to ensure the reliability of the genotype calls.

Mitochondrial and glycolysis stress tests

This assay were performed by myself in Dr. Geoffrey D. Girnun's lab at Stony Brook University. Real-time analysis of mitochondrial respiration and glycolytic flux of intact primary fibroblasts was evaluated by measuring the oxygen consumption rate (OCR, a result of oxidative phosphorylation) and extracellular acidification rate (ECAR, largely due to glycolytic lactic acid production) respectively using Mito an XF-96 Extracellular Flux Analyzer (Seahorse Bioscience, XF Cell Mito Stress Test Kit, # 103015-100, and XF Glycolysis Stress Test Kit, # 103020-100). One day before each assay run, 20,000 primary fibroblasts consisting of 3 WT lines and one S37P line were seeded per well of a XF96 cell culture microplate (23 biological replicates for

each cell group) and incubated for 24 h to allow attachment. XF96 sensor cartridges (Seahorse Bioscience) were hydrated overnight in a non-CO2 incubator.

On the day of glycolysis stress test assay, cells were washed and equilibrated for 1 h at 37 °C in a non-CO2 incubator with unbuffered XF assay medium. For the mito stress test assay, cells were first washed and equilibrated for 1 h at 37 °C in a non-CO2 incubator with unbuffered XF base medium supplemented with 25 mM glucose (Sigma-Alorich, # D9434), 1 mM pyruvate, and 4 mM L-glutamine (Thermo Fisher Scientific, # 25030081). Mitochondrial function was evaluated via sequential injections of oligomycin (1 μM), carbonyl cyanide-4 (trifluoromethoxy)phenylhydrazone (FCCP, 1 μM), and rotenone/antimycin A (0.5 μM). Glycolytic function was examined via sequential injections of glucose (10 mM), oligomycin (1 μM), and 2-Deoxy-D-glucose (2-DG, 50 mM). Optimal cell number, and concentration of oligomycin and FCCP were obtained after a series of titration experiments for each mitochondrial or glycolysis effector to elicit maximal responses without toxicity (data not included in this thesis).

Following each of the above compound injections, there are 3 min of mixing and 3 min of measuring for OCR and ECAR at 37 °C under air. All measurements from above assay runs were normalized to cell number by CyQUANT cell proliferation assay (Thermo Fisher Scientific, # C7026) according to the manufacturer's protocol.

All above experiments under this section were repeated at least three times.

Data analysis

Raw data of seahorse assays was visualized and analyzed using Wave Desktop 2.3, and data outliers due to random errors (i.e. negative values) were removed from the downstream analysis. Seahorse XF Report Generators were then used to calculate individual parameters for further analysis. Seven oxygen consumption variables were calculated as previously described (Dranka, Hill, & Darley-Usmar, 2010) for evaluating mitochondrial function, including basal, non-mitochondrial, ATP linked, non-ATP linked, maximum, reserve capacity, and coupling efficiency. For glycolytic function assessment, four ECAR variables were calculated including non-glycolytic acidification, glycolysis, glycolytic capacity, and glycolytic reserve. GraphPad Prism 6 was used for graph generation.

In brief, the above individual parameters were calculated based on the below equations:

Parameters	Calculation
non-mito resp.	min rate measurement after rotenone/antimycin-A injection
basal resp.	(last rate measurement before first injection) – (non-mito resp. rate)
maximal resp.	(max rate measurement after FCCP injection) – (non-mito resp.)
proton leak	(min rate measurement after oligo injection) – (non-mito resp.)
ATP production	(last rate measurement before oligo injection) – (min rate measurement after oligo injection)
spare respiratory capacity	(max resp.) – (basal resp.)
spare respiratory capacity %	(max resp.) / (basal resp.) × 100
coupling efficiency %	(ATP production rate) / (basal resp. rate) × 100
non-glycolytic acidification	last rate measurement after 2-DG injection
glycolysis	(max rate measurement before oligo injection) – (last rate measurement before glucose injection)
glycolytic capacity	(max rate measurement after oligo injection) – (last rate measurement before glucose injection)
glycolytic reserve	(glycolytic capacity) – (glycolysis)
glycolytic reserve %	(glycolytic capacity rate) / (glycolysis) × 100

Resp., respiration; max, maximum; min, minimum; mito, mitochondrial.

Statistical analysis

All experiments were repeated three times to ensure the reproducibility of the results. All quantitative results are expressed as mean ± standard deviation. Unless otherwise indicated, statistical analysis was done using one-way ANOVA with Tukey-Kramer post-hoc test for a single experiment, and two-way ANOVA with Tukey-Kramer post-hoc test for multiple independent experiments between patient and WT control groups. For mitochondrial and glycolysis stress tests, a linear mixed-effects model was fitted using R-3.3.2 software. *P* values that are below 0.05 were considered significant.

Assay	# of lines used	# of passage	# of replicates per line	# of successful repeats
XTT viability	2 WT, 1 S37P	13	4 wells for each density	4
Growth curve	2 WT, 1 S37P	4, 12	3 wells for each day	2
BrdU proliferation	2 WT, 1 S37P	4, 12	3 coverslips	2
Mitochondrial staining	1 WT, 1 S37P	7	3 coverslips	1
Bioenergetics tests	3 WT, 1 S37P	4, 5, 6	23 wells	3

3.3 Results

Clinical characteristics of Ogden syndrome

Detailed clinical data of Ogden syndrome patients was collected by various physicians, and systematically reviewed, analyzed and organized by myself (Table 3.1). Numbers in brackets in the text below indicate affected/evaluated (with clinical information available) probands.

Preterm birth of Ogden syndrome infants is commonly seen (5/8). The perinatal course is usually complicated with symptoms such as respiratory distress. All eight male infants required a neonatal intensive care unit (NICU) stay, ranging from 10 days to a month. Both pre- and post-

natal growth failure (6/6) is severe with growth parameters usually below 5th centile for weight, length, and head circumference. The mean values for birth weight, length, and head circumference among Ogden syndrome patients are 2.45 kg (< 5^{th}), 46.29 cm (5-10^{th}), and 31.61 cm (< 5^{th}) respectively. Developmental delay (5/5) is severe, sometimes associated with hypotonia and evidence of cerebral atrophy via neuroimaging.

All eight Ogden syndrome patients shared similar craniofacial features (Fig 3.1), which includes prominent eyes, down-slanting palpebral fissures, thickened or lagging eyelids, large ears in relation to body size, flared or prominent nostrils, short columella, protruding upper lip or long philtrum, microretrognathia, small chin (5/6), narrow palate, wrinkled forehead, large fontanels (6/7) and hypertelorism. There is also delayed closing of fontanels. These patients have scalp hair, and their eyebrows are often very fine. They also have redundancy or laxity of the skin (3/6) with minimal subcutaneous fat (6/7), cutaneous capillary malformations (2/6), and often acquired recurrent eczema (5/6).

Cardiac dysfunction is a universal finding in Ogden syndrome patients. 5/7 of Ogden syndrome patients were born with congenital heart defects, including patent ductus arteriosus, and/or patent foramen ovale, and/or ventricular/atrial septal defect, and/or branch pulmonary artery stenosis, and/or pulmonary hypertension (newborn), and/or bicuspid aortic valve, and/or dextroposition. And 5/7 of them developed an enlarged heart and decreased systolic function while alive. Furthermore, cardiac arrhythmias and other ECG abnormalities are universal findings among all affected individuals, including premature atrial contractions (PACs), and/or premature ventricular contractions (PVCs), and/or supra-ventricular tachycardia (SVT), and/or ventricular tachycardia (VT), and/or atrial fibrillation, and/or bradycardia, and/or intraventricular conduction delay, and/or QT prolongation, and/or Torsade de Points (TDP), and/or nonspecific

T-wave abnormalities. Notably, one of the patients also presented fetal SVT, which occurred at 20 weeks of gestation and was controlled with maternal administration of flecainide.

Musculoskeletal findings include short neck (3/7), scoliosis (3/7), pectus excavatum as well as metatarsus valgus, clinodactyly, hip dysplasia, and broad or relatively large halluces (2/6). Neurobehavioral wise, apneic episodes (5/6) and poor feeding are very common (7/8) symptoms, and many have apparent life-threatening events related to aspiration (5/6). More than half have neonatal hypotonia (4/6). Many have cerebral atrophy (4/5) on brain MRI. Microcephaly (5/7) is commonly seen as well. Genitourinary wise, some affected males have inguinal hernia (3/7), cryptorchidism (5/7), and small testes (4/7). Large phallus has also been observed (1/8). Recurrent infections are common (5/5) and can include frequent otitis media, multiple gastrointestinal and respiratory viral infections, post-operative infection, etc. Some sporadic findings among patients can include umbilical hernia, possible hearing loss, glomerulosclerosis, cephalohematoma, and cystic renal dysplasia.

All five female carriers of Ogden syndrome are asymptomatic. However, X-chromosome inactivation studies of B cells of these women revealed skewing of inactivation of the X-chromosome favoring the wild-type NAA10 allele (Myklebust et al., 2014). Dr. Lyon, who collected and interviewed the Utah family also reported that some female carriers might have intellectual disability, although no formal IQ testing was ever done.

Table 3.1 Clinical features of Ogden syndrome

Features		Family 1 (Utah)					Family 2 (California)		
	affected/ evaluated	II-1	II-6	III-4	III-6	III-7	II-1	III-2	III-4
Gestation									
Complications	6/7	late placental insufficiency	late placental insufficiency	gestational diabetes	gestational diabetes	gestational diabetes	NK	fetal SVT	-
Term (w)	preterm 5/8	37[1/2]	38 [1]	37 [3]	35 [4]	33 [2]	43	39 [5]	33
Birth									
C-section	3/7	-	-	-	-	+	+	+	NK
Weight (centile)		2.14 kg, < 5th	3.065 kg, 10th-25th	2.41 kg, < 5th	2.604 kg, 5th-10th	1.559 kg, < 5th	3.3 kg, 25th-50th	2.66 kg, 5th-10th	1.87 kg, <5th
Height (centile)		47 cm, 10th-25th	47 cm, 10th-25th	44 cm, < 5th	48 cm, 10th-25th	39 cm, < 5th	51 cm, 50th-75th	46 cm, 5th-10th	48.3 cm, 25th-50th
Head circumference (centile)		32 cm, < 5th	33.75 cm, 10th-25th	32 cm, < 5th	32.5 cm, 5th-10th	27 cm, < 5th	32 cm, < 5th	32 cm, < 5th	NK
Apgar scores		1[1], 1[5]	4[1], 6[5]	6[1], 7[5], 9[10]	4[1], 9[5]	4[1], 8[5]	6[1], 8[5]	7[1], 9[5]	NK
Perinatal course									
Complications	7/7	meconium aspiration leading to CNS depression	respiratory distress, mild hyperbilirubinemia, poor feeding	apnea, thrombocytopenia, polycythemia, hyperbilirubinemia, feeding difficulties.	respiratory distress, feeding difficulties, mild hyperbilirubinemia	respiratory distress, polycythemia, jaundice, low cortisol, hypotension, feeding and growth difficulties	poor feeding, jaundice, vasomotor instability?	NK	jaundice
NICU stay	8/8	+?	11 days	24 days	2 weeks	5 weeks	+?	10 days	21 d
Hypoglycemia	2/7	-	-	+	+	-	NK	-	-
Growth									
Prenatal onset Growth retardation	5/7	+	+	+	+	+	-	-	NK
Postnatal growth retardation	6/6	+	+	+	+	NK	+	+	NK
Global delays	5/5	+	+	+	NK	NK	+	+	NK
Craniofacial									
Prominent eyes	4/6	+	+	+	+	NK	-	-	NK
Downslanted palpebral fissures	4/7	-	+	+	+	+	-	-	NK
Thickened or lagging eyelids	2/6	-	-	-	-	NK	+	+	NK
Large ears in relation to body size	6/6	+	+	+	+	NK	+	+	NK
Flared or prominent nostrils	3/6	+	+	+	-	NK	-	-	NK
Short columella	6/6	+	+	+	+	NK	+	+	NK
Protruding upper lip or long philtrum	5/6	-	+	+	+	NK	+	+	NK

Micro- or retrognathia	5/7	+	+	+	+	+	-	-	NK
Small chin	5/6	+	+	+	+	+	-	NK	NK
Narrow palate	4/6	+	+	+	+	NK	-	-	NK
Wrinkled forehead	5/5	?	+	+	?	+	+	+	NK
Microcephaly	5/7	+	-	+	?	-	+	+	+
Hypertelorism	3/7	-	-	+	+	+	-	-	NK
Gastroesophageal reflux	3/3	NK	NK	+	+	NK	NK	+	NK
Musculoskeletal									
Large fontanels	6/7	+	+	+	+	+	+	-	NK
Short neck	3/7	+	+	+	-	-	-	-	NK
Broad or relatively large halluces	2/6	+	-	-	+	NK	-	-	NK
Scoliosis	3/7	-	-	+	+	-	+	-	NK
Metatarsus valgus	2/8	+	+	-	-	-	-	-	-
Integument									
Lax or redundant skin	3/6	-	-	+	+	NK	+	-	NK
Minimal or patchy subcutaneous fat	6/7	+	+	+	+	-	+	+	NK
Cutaneous capillary malformation	2/6	-	+	-	-	NK	-	+	NK
Eczema	5/6	+	+	+	+	NK	-	+	NK
Cardiac									
Structural & functional defects	5/7	peripheral PS, a septal defect	PDA	VSD, PFO, trivial PS, trivial bilateral branch pulmonary artery stenosis	ASD/PFO, PDA, mild PHTN, thickened bicuspid aortic valve, dextroposition	PDA, PFO, decreased ventricular systolic functions, an enlarged right ventricle, abnormal appearing aortic valve	-	-	NK
Cardiomegaly	5/7	+	+	+	+	+	-	-	NK
Arrhythmias	8/8	PAC, PVC, SVT, VT	VT, AF	PAC, PVC, TDP, non-specific T wave abnormality, bradycardia	PVC, PAC	VT, bradycardia	SVT (AVNRT)	fetal SVT, VT	+, ICD was implanted
Death associated with apparent cardiogenic Shock +/- arrhythmia	7/8	+	+	+	-	+	+	+	+
Genital									
Inguinal hernia	3/7	-	-	-	+	+	-	+	NK
Cryptorchidism	5/7	+	+	+	+	+	-	-	NK
Small testes	4/7	+	+	-	-	+	+	-	NK
Neurobehavioral									
Hypotonia	3/6	-	-	+	+	NK	-	+	NK
Poor feeding	7/8	+	+	+	+	+	+	-	+

				Nissen fundoplication/G-tube placement, Laparoscopic gastrostomy tube placement	Nissen fundoplication with G-tube placed, Laparoscopic gastrostomy tube placement	Naso/orogastric feeding tube placement			
Dysphagia	3/3	NK	NK	+	+	+	NK	NK	NK
Apneic episodes, ALTE or aspiration	5/6	+	+	+	?	NK	+	-	+
Cerebral atrophy (enlarged ventricles)	4/5	+	+	+	-	NK	NK	+	NK
Hematologic/lymphatic/immunologic									
Recurrent infections	5/5	diarrhea	otitis media, bronchiolitis	viral bronchiolitis, surgical infection	otitis media, viral bronchiolitis	Viral infections	NK	-	NK
Hematologic	4/6	-	iron deficiency anemia	polycythemia	low platelets	polycythemia	NK	-	NK
Other features		clinodactyly	umbilical hernia, questionable hearing, pectus excavatum	glomerulosclerosis, cerebral dysgenesis	hydrocele, cephalohematoma	cystic renal dysplasia	hip dysplasia, large phallus	micro- and macrovesicular steatosis, lagophthalmos	-
Age of death (m)	average 10.2	$11^{1/2}$	$9^{1/2}$	15	$4^{1/2}$	$5^{1/2}$	8	$11^{1/2}$	16
Autopsy	6/6	+	+	+	NK	+	+	+	NK

"+", feature present; "-", feature absent; AF, atrial fibrillation; ASD, atrial septal defect; AVNRT, AV nodal reentrant tachycardia; CNS, central nervous system; ICD, implantable cardioverter defibrillator; NK, not known; PAC, premature atrial contraction; PDA, patent ductus arteriosus; PFO, patent foramen ovale; PHTN, pulmonary hypertension, newborn; PS, pulmonary stenosis; PVC, premature ventricular contraction; SVT, supraventricular tachycardia; TDP, Torsades de pointes; VSD, ventricular septal defect; VT, ventricular tachycardia. "?" indicates length of stay is not known.

Validation of NAA10 S37P allele in patient's fibroblasts

Upon receiving the primary fibroblasts from the one and only available Ogden syndrome patient (Utah family, III-6), I isolated its genomic DNA and validated the NAA10 c.109T>C variant allele by a Sanger sequencing experiment. The WT alleles from the control samples were also evaluated and confirmed (Fig 3.2).

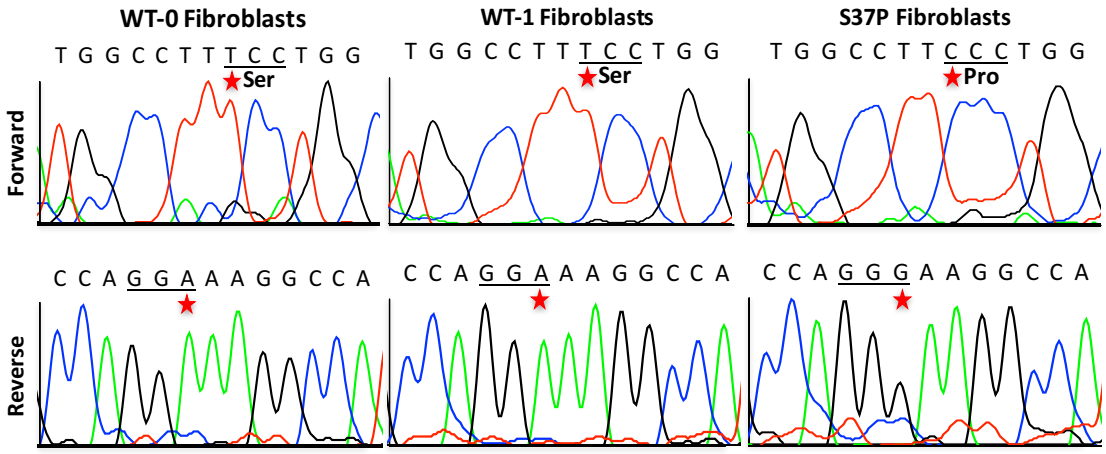


Figure 3.2 Sanger sequencing analysis of NAA10 S37P allele in primary fibroblasts. Sanger sequencing of genomic DNA from Ogden syndrome patient's fibroblasts confirmed the c.109T>C variant in exon2 of NAA10. Sequence information of both forward and reverse strands are shown. The c.109 allele is indicated by the red asterisk, and the codon for p.S37 is highlighted.

NAA10 S37P impairs NatA complex formation

The crystal structure of human NAA10 and NatA has yet to be solved, but based on the crystal structure of the NatA complex from S. pombe (Liszczak et al., 2013), and the fact that proline is a well-known helix breaker and serine 37 is right at the end of helices α2 of NAA10, we hypothesized that the S37P mutant has a negative effect on NatA (NAA10-NAA15) complex formation. To investigate this possibility, I performed a co-immunoprecipitation (co-IP) experiment by co-expressing NAA15-myc and NAA10-V5 or NAA10-S37P-V5 in HEK293 cells in collaboration with Max Doerfel. The result revealed that the interaction between NAA10-S37P-V5 and total NAA15 was significantly reduced compared to that between NAA10-WT-V5 and NAA15 (Fig 3.3). This observation was further validated by detecting the endogenous protein interaction between NAA10 (WT or mutant) and NAA15 in both HEK293 cells and TERT-immortalized fibroblasts from WT and Ogden syndrome patient (Myklebust et al., 2014).

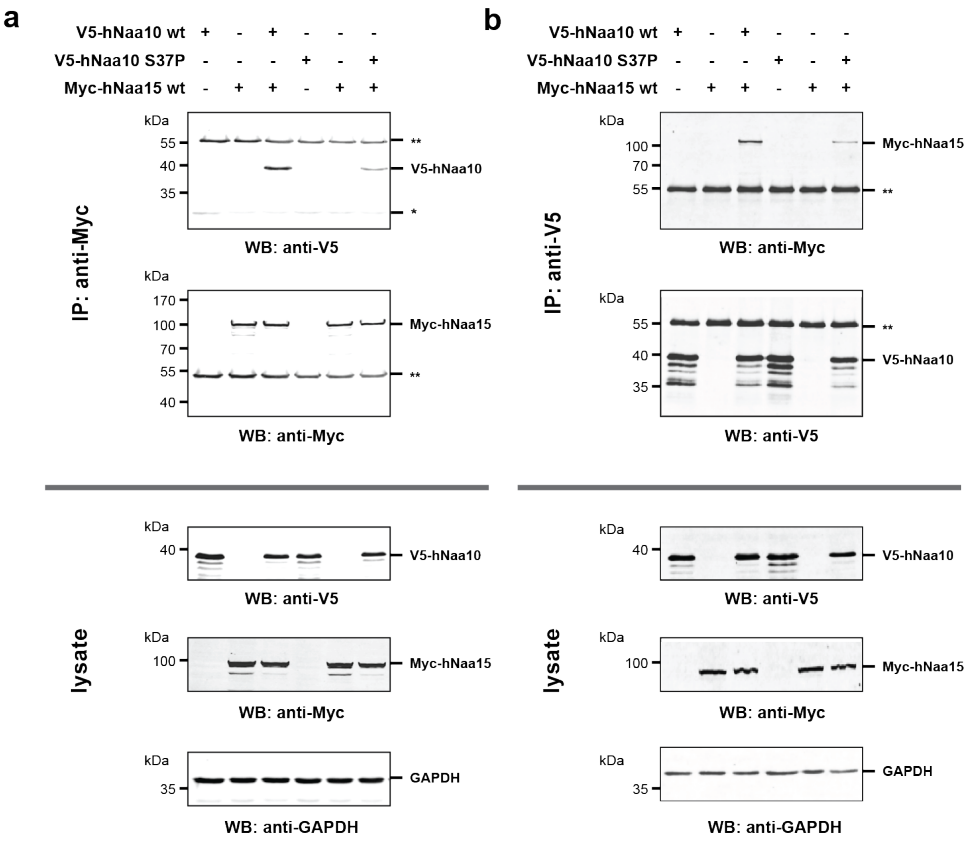


Figure 3.3 Reduced NatA complex formation of NAA10 S37P. HEK293 cells were transiently transfected with NAA10-V5 (WT and S37P) and NAA15-Myc encoding plasmids and NAA15-Myc (a) or NAA10-V5 (b) was immunoprecipitated using anti-V5 or anti-Myc antibodies, respectively (upper panel). Loading and transfection controls are shown in the lower panel. ** and * indicate heavy and light chain of precipitating antibody respectively. GAPDH levels indicate equal input. Max Doerfel performed the experiments in panel (a). Max repeated the experiment in (a) three times, I repeated the experiment in (b) four times.

NAA10 S37P fibroblasts show altered growth and proliferation

When I started to culture primary fibroblasts from the patient and WT under an optimized culture condition, I noticed the patient's fibroblasts were growing quite slowly compared to the control. Hence, I first performed a growth curve analysis to assess the growth behavior of S37P fibroblasts. In order to eliminate the confounding factors introduced by senescence and/or other things that possibly occur during long-term culture (e.g. spontaneous mutations), I chose to use cells with a low passage number (P4) for this experiment. 20K fibroblasts per well were seeded in 12-well plates for this growth curve experiment. I chose to plate 20K cells per well because I read several papers with the reports of similar experiments, and this number seems to be a reasonable one for this type of assay. As shown in Fig 3.4 a, primary fibroblasts were cultured and measured over eight consecutive days, and S37P cells showed a significantly decreased growth rate compared to the WT control. Furthermore, a BrdU cell proliferation assay was later carried out using fibroblasts of the same passage number, and the result confirmed that S37P cells have an overall much slower proliferation capacity compared to the WT control cells (Fig 3.4 b, c). This growth defect was also seen independently by our collaborators when culturing S37P TERT-immortalized fibroblasts compared to their WT TERT treated control cells (Myklebust et al., 2014). We are aware that since neither of our controls are isogenic, the differences seen among these cell lines could as well be contributed by factors such as different genetic background.

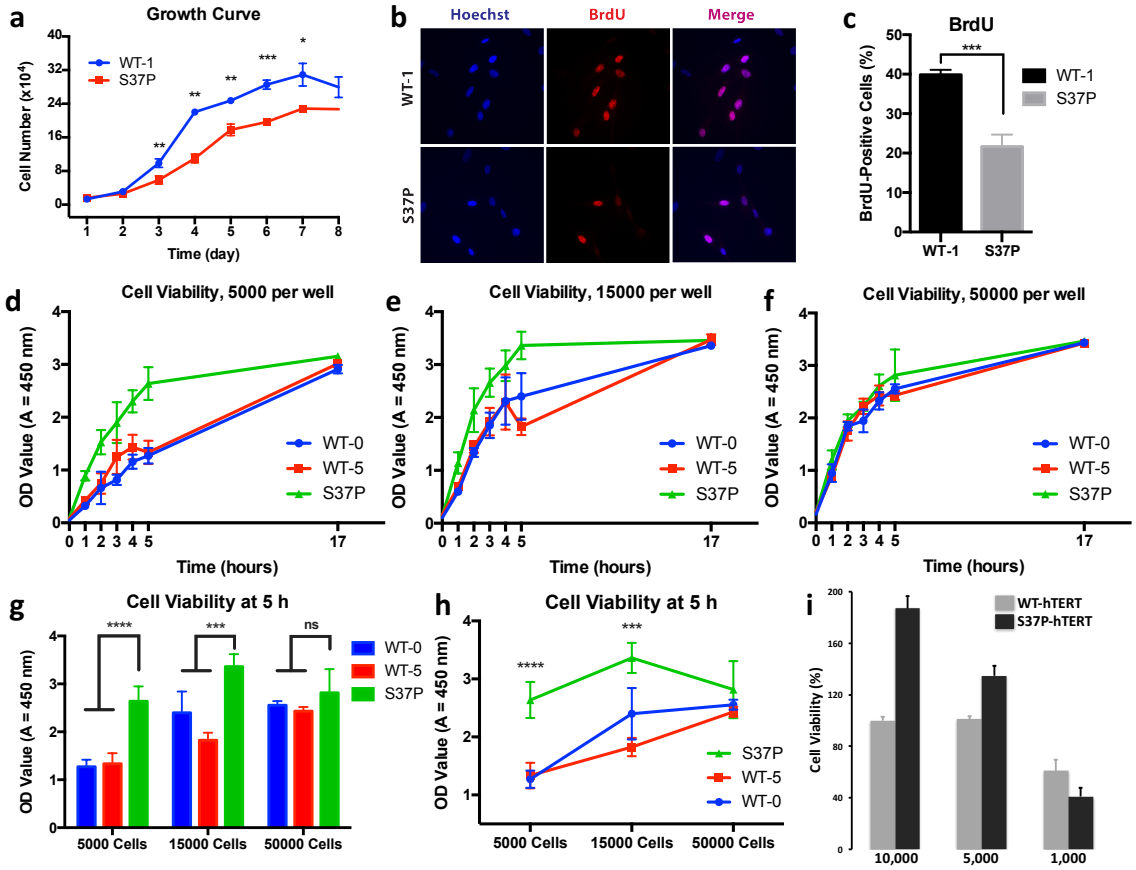


Figure 3.4 Altered growth rate, proliferation and metabolic activity of NAA10 S37P fibroblasts. (a) Growth-curve analysis of the indicated primary fibroblast types. 20K per well of cells were seeded in 12-well plates for this experiment. Statistical analysis was done using two-tailed Student's *t*-test. (b) Representative fluorescence pictures of BrdU incorporation of primary fibroblasts at passage 4. (c) Quantitation of BrdU-positive fibroblasts. Statistical analysis was done using *Pearson's* chi-squared test. (d-f) 2,3-bis-(2-methoxy-4-nitro-5-sulfophenyl)-2H-tetrazolium-5-carboxanilide (XTT) assays showing growth rates of the indicated primary fibroblast types. (g) Bar and (h) line charts of OD measurements of the indicated cells in d-f at 5 h post XTT incubation. (i) WT- and S37P- TERT treated fibroblasts were seeded at 1000, 5000 or 10,000 cells per well in a 96-well plate and incubated for 16 h before being subjected to cell viability/metabolic activity assay (WST-8). All figures are representative data. All data are shown as the mean ± standard deviation. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$; ns, statistically non-significant; OD, optical density. figure (i) was a figure re-used from (Myklebust et al., 2014) with permission. Experiments shown in (d-h) were performed using four different brands of serum. Experiments shown in (a-c) were repeated at least twice.

NAA10 S37P fibroblasts have altered metabolic activities

During a serum test to optimize the fibroblast culture condition, I noticed NAA10 S37P fibroblasts seem to have an unusual metabolic activity pattern responding to different cultural densities compared to the WTs regardless of the serum brands tested (Fig 3.4 d-h). This was evaluated by the XTT assay, a colorimetric assay that assesses cell viability as a function of cell number based on metabolic activity (the pyridine nucleotide redox status of cells) (Berridge et al., 2005; Marshall et al., 1995). Unlike the WT control fibroblasts, in which an overall improved cell viability (proliferation) was seen from a sparse to a dense culture, S37P fibroblasts seem to have the highest metabolic activity in a median dense (15,000/per well in a 96-well plate) culture, whereas there are much less viable cells seen in either a dense (50,000/per well in a 96-well plate) or a sparse culture (5,000/per well in a 96-well plate). When kept in a very dense culture (50,000/per well in a 96-well plate), S37P fibroblasts tend to have indistinguishable cell viability from WT cells. However, when kept in a relatively sparse culture (5,000/per well), S37P fibroblasts showed a significantly more active metabolism compared to the WT controls. A separate WST-8 cell proliferation assay performed later using human telomerase reverse transcriptase (hTERT) immortalized fibroblasts from S37P patient and WT control also demonstrated a similar change of cellular metabolic activity in S37P fibroblasts responding to a varying cultural density (Fig 3.4 i) (Myklebust et al., 2014).

One important aspect of cell metabolism is to generate energy through mitochondrial respiration and glycolysis. An impaired bioenergetics and mitochondrial function has been linked to the pathogenesis of many human diseases especially heart diseases (El-Hattab & Scaglia, 2016; Lopaschuk, 2016), since the heart is one of the most energy demanding organs in the body. The unusual metabolic behavior of S37P cells together with the fact that Ogden syndrome patients

present with heart-related issues prompted me to investigate whether S37P cells have an abnormal bioenergetics pathway.

I first evaluated the mitochondria morphology in S37P fibroblasts using MitoTracker staining. The overall mitochondria structural of patient's cells did not reveal any obvious abnormality compared to the WT controls, with some normal tubular cristae also seen (Fig 3.5). Quantifications of the size and number of mitochondria in the fibroblasts were not performed, therefore, I cannot comment on these aspects. However, the autopsy analysis of the skeletal muscle and heart cells from this patient described normal findings in size, number and appearance of his mitochondria.

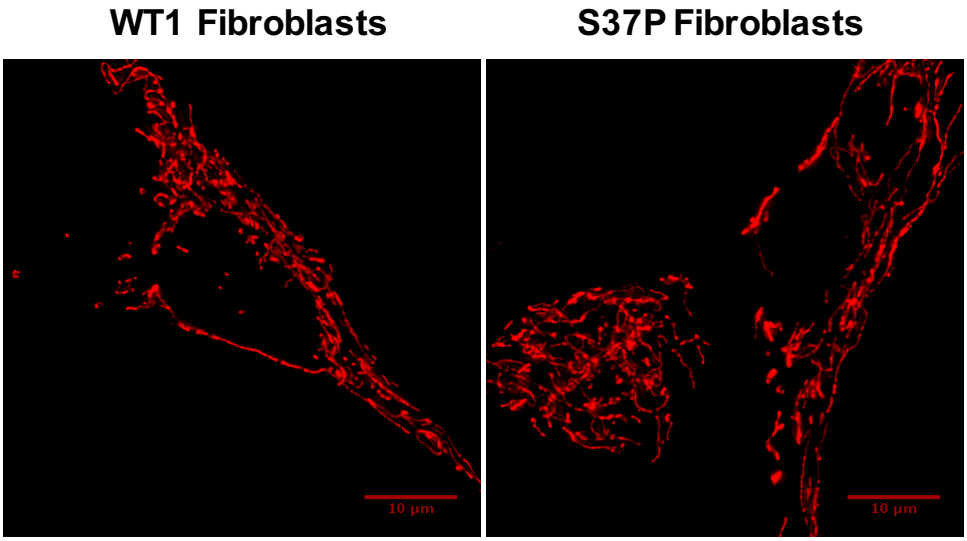


Figure 3.5 Mitochondrial structure evaluation of S37P primary fibroblasts. Representative images of P7 WT1 control fibroblasts and S37P patient fibroblasts stained with MitoTracker Red CMXRos. At least 6 fields of view (FOV) of WT1 and 15 FOVs for S37P cells were surveyed and reviewed for a comparison purpose in two independent experiments. Scale bars, 10 µm.

Next, I examined the mitochondrial function of S37P fibroblasts using XF96 Extracellular Flux Analyzer (Seahorse Bioscience). This analyzer measures the two major energy producing pathways of the cell in real-time, mitochondrial respiration (i.e. oxidative phosphorylation, OXPHOS) and glycolysis. Mitochondria consume oxygen when oxidizing substrates (e.g. fatty acids) to generate ATP, hence, the Seahorse XF Analyzer evaluates the mitochondrial respiration process through measuring the Oxygen Consumption Rate (OCR) of cells. Besides OXPHOS, cells can also generate ATP via glycolysis, which produces lactate and protons and leads to the acidification of the medium surrounding the cells (Brand, 1990). The Seahorse XF Analyzer evaluates the glycolysis process through measuring the Extracellular Acidification Rate (ECAR) of cells.

For mitochondrial function evaluation, primary fibroblasts were cultured under glucose-based medium, where cells can support ATP production mainly through OXPHOS. After serially injecting oligomycin (ATP synthase inhibitor), FCCP (uncoupler of OXPHOS), and rotenone/antimycin A (inhibitors of electron transport chain complex I and III, respectively), key parameters of metabolic function were revealed, including basal respiration, ATP production, proton leak, maximal respiration, spare respiratory capacity, non-mitochondrial respiration and coupling efficiency (detailed explanation for the calculation of individual parameters can be found in the table under Data Analysis). S37P fibroblasts showed a very similar mitochondrial function compared to the WT control fibroblasts under this *in vitro* condition, with all OCR parameters evaluated closely resembling those in the WT control fibroblasts (Fig 3.6 a, c). However, a significant elevation of ECAR in S37P fibroblasts during the whole assay run was noticed (Fig 3.6 b). Since the main contributor of ECAR in this assay is the production and extrusion of protons into the extracellular medium during glycolysis, my next interest was to

examine whether the elevation of ECAR seen here was due to an altered glycolytic function in S37P cells.

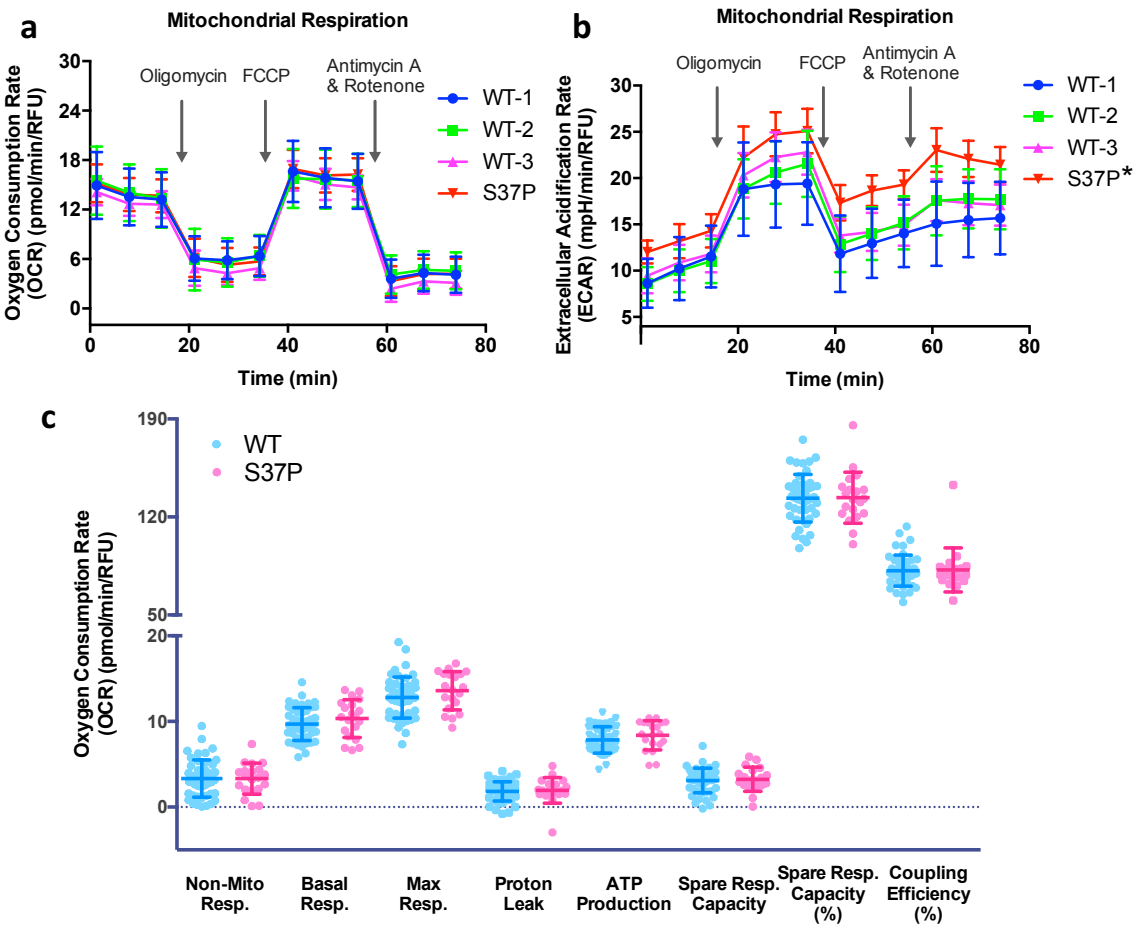


Figure 3.6 Real-time mitochondrial respiration assessment of NAA10 S37P fibroblasts. Mitochondrial respiration (oxygen consumption rate, OCR) analysis was performed using XF-96 Extracellular Flux Analyzer (Seahorse Bioscience) as described in METHODS. Representative graphs of OCR (a) and ECAR (b) output from XF96 analyzer of primary fibroblasts and their responses to oligomycin, FCCP, and rotenone/antimycin A. The arrow indicates the injection of each compound. *$p <$ 0.0001 ($p = 0.0001$ for measurement 6) compared pooled WT and S37P at each measurement by Student's *t*-test. (c) Representative eight individual oxygen consumption variables representing mitochondrial function were calculated based on equations described in the table under Data Analysis. WT is a pool of each individual parameter calculated from WT-1, 2, and 3 controls. # of data points (replicates) in figure (c), WT: 47, S37P: 20. Results are shown as the mean ± standard deviation. Statistical analysis was done using a linear mixed-effects model. This experiment was repeated three times.

For glycolytic function evaluation, cells were cultured in glucose-free assay medium, and glycolytic flux was measured by assessing the extracellular acidification rate (ECAR) after a sequential injection of glucose, oligomycin, and 2-deoxyglucose (2-DG). The addition of oligomycin forces the cells to produce ATP via glycolysis to maintain energy homeostasis (Wu et al., 2007), thus it provides a measure of maximum glycolytic potential. 2-DG is a glucose analog that inhibits glycolysis through competitive binding to glucose hexokinase, the first enzyme in the glycolytic pathway. Non-glycolytic acidification, which is caused by processes in the cell that produce protons other than glycolysis such as the tricarboxylic acid (TCA) cycle and glycogenolysis (Ipata & Balestri, 2012), was defined as the ECAR after the treatment of 2-DG (V. Chen et al., 2012; Hill et al., 2012). S37P fibroblasts grown under this condition showed similar non-glycolytic acidification compared to the controls, as well as their glycolytic capacity and glycolytic reserve (Fig 3.7 d). However, a significant increase in the ECAR associated with glycolysis was seen in S37P cells (Fig 3.7 a, c). This result offered an explanation of the elevated ECAR seen in the mitochondrial stress test assay above. In the meantime, comparable OCR values during the process between the patient cells and the WT controls further validated an unchanged mitochondrial function of S37P cells seen in the above assay (Fig 3.7 b). (Detailed explanation for the calculation of individual parameters can be found in the table under Data Analysis)

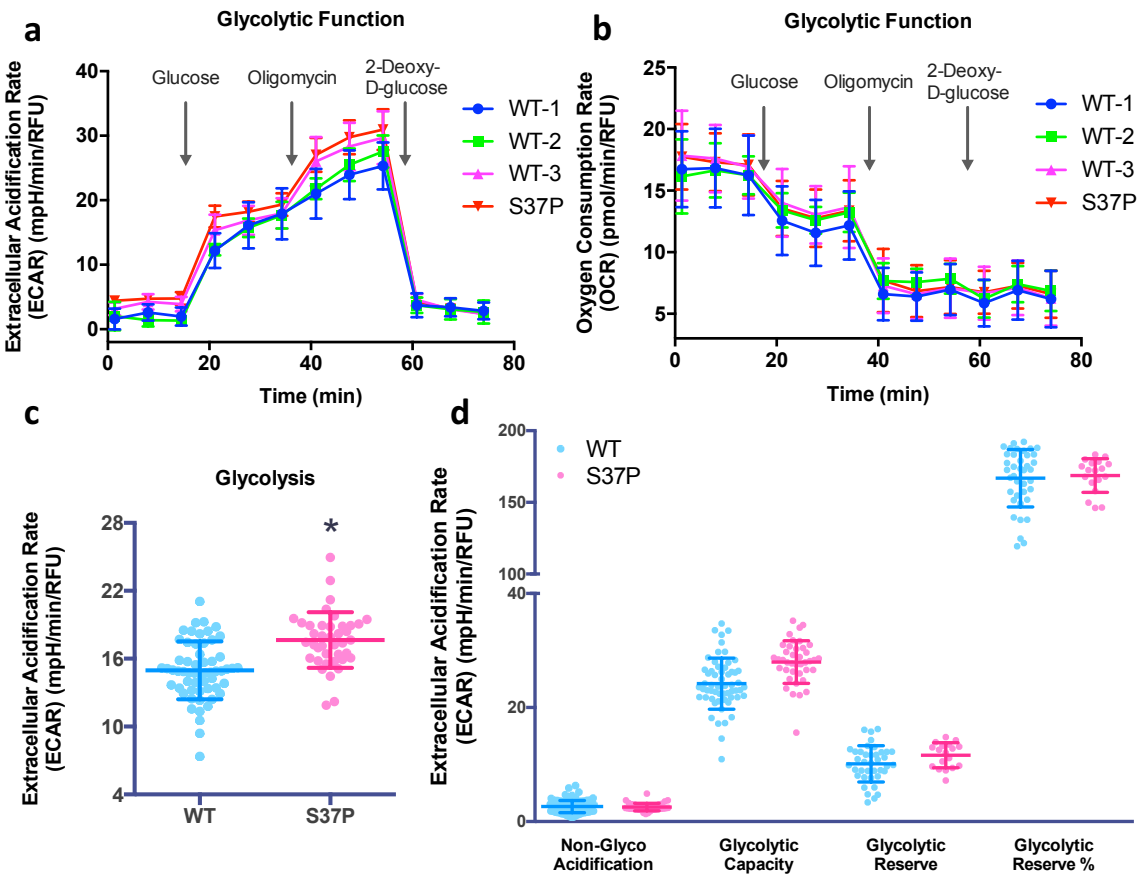


Figure 3.7 Real-time glycolytic activity assessment of NAA10 S37P fibroblasts. Glycolytic rate was measured as extracellular acidification rate (ECAR) using XF-96 Extracellular Flux Analyzer (Seahorse Bioscience) as described in METHODS. Representative graphs of ECAR (a) and OCR (b) output from XF96 analyzer of primary fibroblasts and their responses to glucose, oligomycin, and 2-Deoxy-D-glucose (2-DG). The arrow indicates the injection of each compound. (c) Comparison between S37P and pooled WT control fibroblasts in their glycolysis from all three independent experiments. *$p = 0.018$ by a linear mixed-effects model. # of data points, replicates plus repeats, WT: 58, S37P: 40. (d) Calculation of the four other indicated parameters of glycolytic function in S37P and WT control fibroblasts. WT represented a pool of each individual parameter calculated from WT-1, 2, and 3 controls. Detailed explanation for the calculation of individual parameters can be found in the table under Data Analysis. # of data points, replicates plus repeats, WT: 87 (non-glyco acidification), 58 (glycolytic capacity), 39 (glycolytic reserve); S37P: 40 (non-glyco acidification, glycolytic capacity), 18 (glycolytic reserve). Five parameters of glycolytic function shown in figure (c) and (d) were calculated based on equations described in METHODS. All results are shown as the mean ± standard deviation. Statistical analysis was done using a linear mixed-effects model. This experiment was repeated three times.

3.4 Discussion

Human NTA-deficiency phenotype

Since the discovery of Ogden syndrome in 2011, there has been at least 30 more individuals reported worldwide carrying pathogenic variants in NAA10, among whom 10 are males and 17 are females (Casey et al., 2015; Esmailpour et al., 2014; Popp et al., 2015; Rope et al., 2011; Saunier et al., 2016; Thevenon et al., 2016; and personal communications). One individual not yet reported in the literature is a male infant carrying exact the same c.109T>C variant in NAA10 that caused Ogden syndrome, and he died in infancy too (6 month of age). We recommended to refer to these cases as NAA10-related disorder.

The NAA10 pathogenic variants identified to date are largely missense ones, including p.Ser37Pro (c.109T>C), p.Tyr43Ser (c.128A>C), p. Ile72Thr (c.215T>C), p.Arg83Cys (c.247C>T), p.Val107Phe (c.319G>T), p.Arg116Trp (c.346C>T), p.Phe128Leu (c.384T>A), and p.Phe128Ile (c.382T>A), with only one exception that is a splicing variant, c.471+2T→A (Fig 3.8). Patients with different variants in NAA10 present a wide spectrum of phenotype, especially those with the splicing variant c.471+2T→A who seem to have unique features that none of the others share, such as microphthalmia and/or clinical anophthalmia. In fact, this splicing variant is believed to be associated with Lenz microphthalmia syndrome due to a knock-out effect of WT NAA10, dysregulating the retinoic acid signaling pathway in patients cells (Esmailpour et al., 2014).

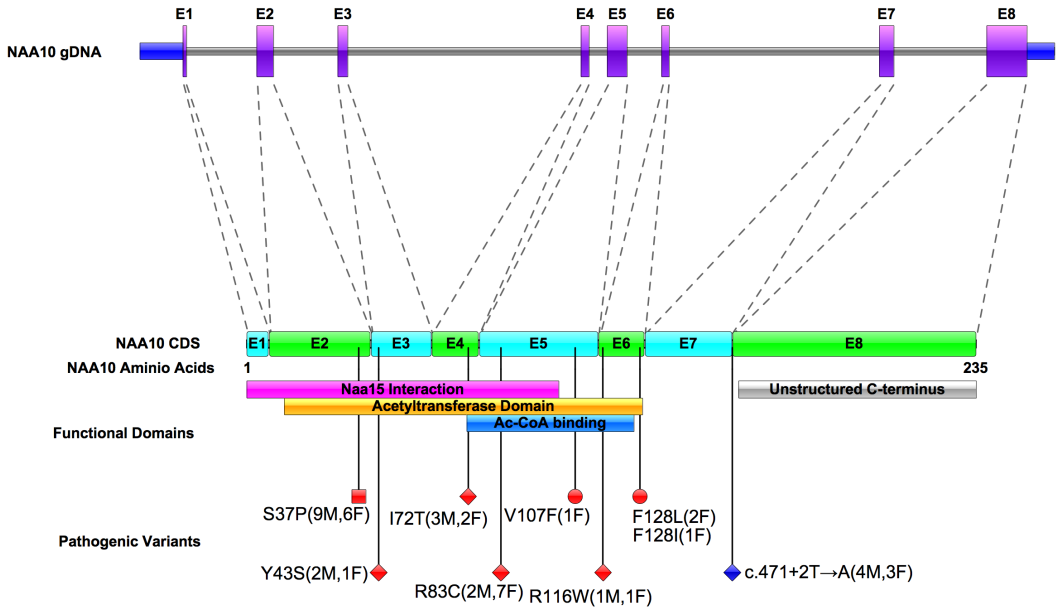


Figure 3.8 Human pathogenic mutations identified in NAA10. A schematic representation of the NAA10 genomic loci, CDS (encoding exons), protein, functional domains (based on NCBI reference NP_003482.1, PDB 4KVX and the crystal structure of the NatA complex (Liszczak et al., 2013) and localization of mutations in affected males (red squares), females (red circles) and a mix of genders (red diamonds). Previously published mutations (numbered after NM_003491.3, NP_003482.1, hg19) are indicated in brackets with the number of affected probands and the specific gender information (M for male, and F for female). The splicing mutation c.471+2T>A was labeled with blue diamonds. Figure was drawn using Illustrator for Biological Sequences (IBS) software version 1.0.2. All structures were drawn to scale.

Despite the disease heterogeneity due to the variant location, genetic and environmental background, all patients with a NAA10-related variant seem to share some common features including developmental delay, intellectual disability, pre-/postnatal growth deficiency, feeding difficulties, dysmorphic features, hypotonia, cardiac dysfunction, skeletal anomalies, and autistic-like behaviors to various degrees. These symptoms are also shared among patients carrying pathogenic NAA15 variants (Stessman et al., 2017; Zaidi et al., 2013; and personal communications), hence they seem to represent a general consequence of a disrupted NatA-mediated Nt-acetylation in humans. A NAA10 knockout mice model (the only NAA10 mice model published) also reported some bone abnormalities, although the phenotype seen in these neonatal mice does not seem to be dramatic (Yoon et al., 2014). However unfortunately, the authors did not report any other phenotypic data associated with their NAA10 knockout mice in this paper.

All NAA10 pathogenic variants reported in humans to date lead to various degrees of impairment in the enzymatic activity of NatA *in vitro*, which was believed to be the main reason for the patients' clinical manifestation. However, of note, the disease onset and the phenotypic severity does not seem to correlate with the remaining levels of the NatA enzyme. For example, both Y43S (in two males) and V107F (in a female) variants leads to a significant reduction (85% and 95% respectively) of NatA *in vitro* activity, with Y43S mutant being even less stable compared to S37P which only leads to 30~70% *in vitro* reduction (depending of the substrate oligopeptide); however, S37P causes lethality in infancy whereas patients with Y43S and V107F have relatively minor clinical features (Casey et al., 2015; Popp et al., 2015).

Male patients with NAA10-related disorders in general have much more severe phenotype than females, with many dying in infancy. One reason behind this observation is that

the females are heterozygous and have X-chromosome inactivation (XCI) process (LYON, 1961; Yang et al., 2011), which could serve as a protective mechanism if the X chromosome carrying the mutation is underrepresented (skewed), as seen in four asymptomatic female carriers of Ogden syndrome as well as female carriers in many other X-linked disorders (Eble et al., 2009; Hedera & Gorski, 2003; Knudsen et al., 2006; Viggiano et al., 2013; Wimplinger et al., 2007). Whereas in males, the mutation is hemizygous, so there is no WT allele to compensate in any way. Notably, NAA10 is not one of those 15% of X-linked genes that escape XCI (Carrel & Willard, 2005). However, the girl with V107F variant has a random XCI pattern, with both the disease allele and WT allele equally expressed in her blood cells. In fact, XCI pattern was random in half of tested females carrying variants associated with NAA10-related disorders (6/12), so the reason for differences in phenotype in female remains unclear (and note that the severity of phenotype in females also varies even when they carry the same mutation). One possible explanation is that female carriers may have a skewed XCI pattern in their cells other than blood (the tested cell type) which gives them an overall less severe phenotype. However, as seen in many other X-linked disorders such as Rett syndrome, a paradigm for XCI studies, any convincing correlation among the XCI pattern of female patients, the presence of X-linked gene mutation, and the severity of the phenotype is still yet to be elucidated (Vacca et al., 2016). And like NAA10-related disorder, Rett syndrome also exhibits substantial phenotypic variability.

Molecular mechanistic paths to Ogden syndrome

In the past couple of years, much research has demonstrated the important role of NAA10 or NatA (NAA10 and NAA15) in cancer (Dörfel & Lyon, 2015; Drazic et al., 2016; Kalvik & Arnesen, 2013), but the evidence is mixed. Some claim NAA10 has tumor suppressor properties (Hua et al., 2011; Kuo et al., 2010; Shin et al., 2009), and others recognize it as a pro-

proliferative and anti-apoptotic protein where an overexpression correlates to a low survival rate and aggressiveness of tumors (Lee et al., 2010; Ren et al., 2008). This contradiction is not surprising if one considers the fact that the substrates of NatA are so diverse and vast in the human proteome, and that NAA10 can also function independently (and post-translationally). Nonetheless, it is clear that NAA10 can regulate cell survival, proliferation and metabolism in cancer cells via several different pathways involving the regulation of cyclin D1, Beclin 1, P53, TSC2, DNA methyltransferase 1 (DNMT1), etc. (Fisher et al., 2005; Funderburk et al., 2010; Gromyko et al., 2010; Lee et al., 2010; Kuo et al., 2010; Lim et al., 2006). The fact that we observed S37P cells have an altered growth rate, proliferation pattern and unusual metabolic activity change responding to different densities might be related to the regulation of NAA10/NatA in some of these signaling pathways. In fact, our collaborators later showed that S37P fibroblasts have an increased expression of Retinoblastoma 1 (RB1) (Myklebust et al., 2014), which is known to be a negative regulator of the cell cycle (Du & Pogoriler, 2006; Manning & Dyson, 2011; Nevins, 2001). It needs to be pointed out that RB1 itself does not seem to subject to NTA modification at all, therefore, NAA10 S37P might have an influence on its expression through a NatA-independent pathway. They also showed that a fraction (at least 9%) of the S37P-hTERT cells did not enter the G0 phase when kept at a high density whereas WT cells were predominantly (over 96%) in the G0 phase (Myklebust et al., 2014). This observation is in line with an earlier yeast study in which it was shown NAA10 deficient cells fail to enter stationary phase, were defective in responding to the mating pheromone α-factor, and were unable to sporulate (Whiteway & Szostak, 1985). This failure to enter the stationary phase of S37P cells under certain situations might contribute to many developmental defects seen in

NAA10-related disorder patients, since cell differentiation and tissue organization is known to require strict control of cell division.

The observation of S37P fibroblasts upregulating their glucose metabolism pathways through glycolysis is interesting. An increased uptake of glucose and the ability to drive glycolysis in oxygen-rich environments is a well-known signature of malignant and rapidly-growing tumor cells termed the Warburg Effect (Warburg, 1956). The elevation of glycolysis seen in S37P fibroblasts partially resemble the Warburg Effect, although the scale of the upregulation of glycolysis detected in S37P fibroblasts is not as dramatic as that seen in cancers. However, since skin fibroblasts we studied are not metabolically active tissues, it is possible that the alteration seen in these cells will be much more amplified in other cell types such as cardiomyocytes where metabolism is more active and energy is more demanding. Hence, it will be really informative to repeat the bioenergetics characterization assays using patient's heart cells to help connect this metabolic change with the patient's phenotype. Moreover, the upregulation of glycolysis in cancers is usually paired with a significantly downregulated mitochondrial respiration, which is also not the case in S37P fibroblasts. The fact S37P fibroblasts show an elevated glycolysis in the circumstance of a normal mitochondrial respiration might indicate a higher energy demand in these cells owing to a less efficient ATP utilization or too much energy wasting at the cellular level. There are indeed a few evidences to date suggesting NAA10p mediated Nt-acetylation can be regulated by the metabolic state of the cell, such as the level of the metabolite Ac-CoA (Yi et al., 2011), and through being regulated by hypoxia inducible factor-1 (HIF-1) in cancers (Lim et al., 2008); however, an active role of NAA10/NatA in regulating metabolism is yet to be discovered. Our bioenergetics

characterization of S37P fibroblasts might provide the first evidence that such regulation might exist.

Meanwhile, researchers including myself also tried to determine whether the NAA10-S37P mutant had an impact on NatA subunit steady-state protein levels. The protein profile result from my own Western blotting experiment using primary fibroblasts from the patient (Utah family, III-6) demonstrated a mild reduction of NAA10 but not NAA15 expression in the S37P fibroblasts. But since the experiments were not repeated more than twice (with NAA10 only successfully being detected once), I did not feel confident to include these data in the result section. However, the same experiments were also performed by our collaborators using EBV-immortalized B cells and hTERT-immortalized fibroblasts from the same Ogden syndrome family, and the results demonstrated a minor mutation-correlated reduction in the steady-state protein levels of both NAA10 and NAA15 (Myklebust et al., 2014). The primary B cells used in this study were isolated from 11 members of Ogden syndrome family including two S37P males and four female heterozygous carriers, and primary fibroblasts were isolated from the same patient I used. Interestingly, the result of EBV-immortalized B cells also revealed a similar degree of reduction for both proteins in half of the female carriers (2 out of 4). I later compared the protein reduction pattern seen in these females with their individual X-chromosome inactivation (XCI) result which were also collected from their B cells, but did not see any correlation between the protein expression amount and the degree of their XCI skewing. Therefore, the reason for this observation remains unknown. It seems like a mutation-related phenotype at the cellular level in some of the Ogden syndrome carriers is a possibility if this is a reproducible finding. Otherwise, it can be just due to a technic error during the experiment.

Our observation of the decreased NatA complex formation of NAA10 S37P mutant and NAA15 validated a simulation model result from earlier, which was constructed to evaluate the structural effects of the NAA10-S37P variant of human NatA complex with either WT NAA10 or the S37P mutant based on the crystal structure of the NatA complex from S. pombe (Liszczak et al., 2013). The human and S. pombe share 66% identity and 81% similarity of NAA10 sequence, and 39% identity and 57% similarity for NAA15 sequence. The resulting models show that helices α1 and α2 of NAA10 are directly involved in mediating the interaction with its auxiliary partner NAA15, and serine 37 is right at the end of helices α2. It has shown that the S37P mutant shortens helix α2 of NAA10, decreases the atomic fluctuations of helix α1 and loop α1-α2 of NAA10, and also increases the flexibility of loop β2-β3 (Myklebust et al., 2014). Moreover, a S. cerevisiae model of Ogden syndrome was recently established by introducing human wild-type or mutant NatA complex into a NatA-null strain (Van Damme et al., 2014), and was reported in the same paper that there was reduced NatA complex formation when co-expressing NAA10 S37P and NAA15 in the yeast. This evidence supports the notion that the reduction of NatA enzymatic activity seen in S37P mutant was mainly, if not entirely, contributed by the structural effect of NAA10 S37P by mediating a less efficient physical interaction with its auxiliary partner NAA15 in the NatA complex. This result also inspired me to think that less NatA complex formation will mean more free forms of NAA10 and NAA15 proteins in the cells, hence, it is possible (in my mind) that some of the phenotype seen in Ogden syndrome patients might also due to amplified roles of complex-free NAA10 and NAA15 in the cells. Both NAA10 and NAA15 have been shown to function independent of their acetyltransferase activity (Linē et al., 2002; D. T. Martin et al., 2007; Shin et al., 2009; Zeng et

al., 2014), and the difference in their tissue-specific expression patterns is also evidence of this notion (Fig 3.9).

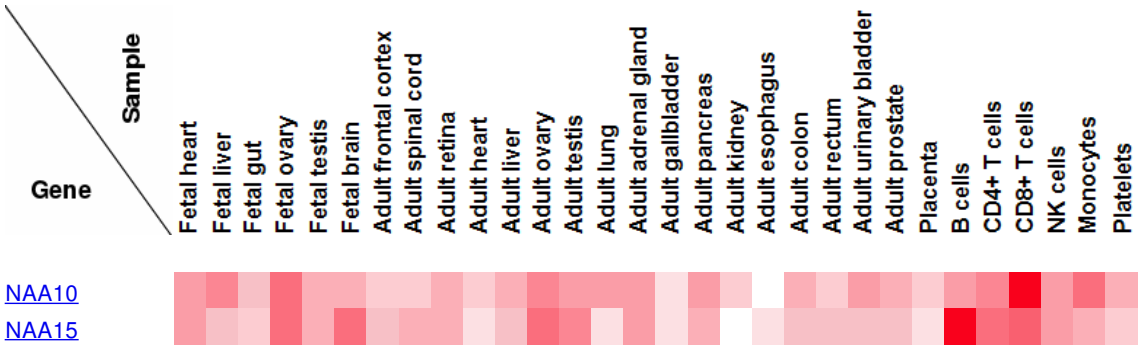


Figure 3.9 Tissue-specific expression pattern of NAA10 and NAA15 in histologically normal human samples. The protein expression of NAA10 and NAA15 in various human tissues was explored based on mass spectrometry-based proteomics data (M.-S. Kim et al., 2014). The figure was generated automatically by Human Proteome Map website (http://www.humanproteomemap.org/).

Furthermore, the yeast model of Ogden syndrome mentioned above also demonstrates that yeast cells expressing human NatA (hNatA) displayed a Nt-acetylation reduction of a large group of NatA substrates when compared to yeast expressing wild-type hNatA. This is in line with a later study of protein N-terminal acetylome in S37P fibroblasts and B cells showing a reduced N-terminal acetylation of a subset of classic NatA substrates in comparison to the WT cells (3% of all N termini (30 out of 1066 N termini)) (Table 3.2) (Myklebust et al., 2014). Interestingly, several proteins that play important roles in protein post-translational modification were identified on this list, such as histone-lysine N-methyltransferase EHMT1, E3 ubiquitin-protein ligase RNF5, 26S protease regulatory subunit 8, dual specificity protein phosphatase 4, deoxyhypusine hydroxylase. This seems to extend the evidence of an earlier discussion on the potential interplay of Nt-acetylation with other protein modifications (Scaglione et al., 2013; Schiza et al., 2013; Tatham et al., 2013). Meanwhile, I also noticed a few important proteins for metabolism from the list such as Ras-related protein RAB32, 39S ribosomal protein L15 (mitochondrial), Nuclease EXOG (mitochondrial), calcium-binding mitochondrial carrier protein Aralar1 (AGC1), etc. which may relate to the bioenergetics alteration we saw in S37P fibroblasts.

Table 3.2 Overview of protein N termini less acetylated in S37P cells and siNatA HeLa cells

P1'[a]	P1-P5[b]	S37P B cells	S37P Fib	siNatA (HeLa)	Description
M	AAADA	√	N.I	N.I	Histone-lysine N-methyltransferase EHMT1 (EHMT1)
M	AAAEE	√	√	N.I	E3 ubiquitin-protein ligase RNF5 (RNF5)
M	AAAQE	√	√	N.I	Cleft lip and palate transmembrane protein 1 (CLPTM1)
M	AADTQ	√	√	√	Translational activator GCN1 (GCN1L)
M	AAESA	√	√	√	Transcription elongation factor B polypeptide 3 (ELOA1)
M	AAGGG	√	√	N.I	Regulation of nuclear pre-mRNA domain-containing protein 2 (RPRD2)
M	AEVQV	√	√	√	60S ribosomal protein L13a (RL13A)
M	AGGGA	N.I.	√	N.I	Ras-related protein Rab-32 (RAB32)
M	AGPLQ	√	√	N.I	39S ribosomal protein L15, mitochondrial (MRPL15/RM15)
M	AIKSI	N.I.	√	N.I	Nuclease EXOG, mitochondrial (EXOG)
M	ALDGP	√	√	N.I	26S protease regulatory subunit 8 (PRS8)
M	AQPGT	√	N.I	N.I	Serum response factor-binding protein 1 (SRFB1)
M	ASAGS	N.I.	√	N.I	Transmembrane protein 47 (TMM47)
M	AVEDE	√	√	N.I	Ras GTPase-activating protein 3 (RASA3)
M	AVFAD	√	√	√	Ribonuclease P protein subunit p30 (RPP30)
M	AVKVQ	N.I.	√	N.I	Calcium-binding mitochondrial carrier protein Aralar1 (AGC1)
M	GAAAA	N.I.	√	N.I	RNA-binding protein 7 (RBM7)
M	GAVTD	√	√	N.I	THO complex subunit 7 homolog (THOC7)
M	GEQPI	√	N.I	N.I	Homer protein homolog 1 (HOME1)
M	SASVV	√	√	N.I	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1 (DAD1)
M	SEGDS	N.I.	√	N.I	Protein RER1 (RER1)
M	SGFLE	√	√	N.I	Transmembrane protein 50A (TM50A)
M	TAQGG	√	√	N.I	Transmembrane protein 85 (TMM85)
M	TKAGS	√	√	N.I	Leucine-rich repeat-containing protein 59 (LRC59)
M	TKVAE	√	N.I	N.I	Probable ATP-dependent RNA helicase DDX10 (DDX10)
M	TMDAL	√	N.I	N.I	Ankyrin repeat and LEM domain-containing protein 2 (ANKL2)

M	TMDKS	√	√	N.I	14-3-3 protein beta/alpha (1433B)
M	VEKEE	√	√	√	SUMO-activating enzyme subunit 1 (SAE1)
M	VEQGD	√	N.I	N.I	Protein LCHN (LCHN)
M	VEYVL	√	N.I	N.I	Pumilio homolog 2 (PUM2)
M	VNPTV	√	√	√	Peptidyl-prolyl cis-trans isomerase A (PPIA)
M	VTMEE	√	N.I	N.I	Dual specificity protein phosphatase 4 (DUS4)

Fib, fibroblasts; a, iMet cleaved off prior to Nt-acetylation; b, first five amino acids of the acetylated peptide; NI, respective N termini were not identified. Table was re-used from (Myklebust et al., 2014) with permission.

Limitations of this study and future directions

Some of the challenges of this study are common to rare disease research field, including the lack of patient samples to increase the statistical power (we only have one patient's sample to use since all other seven patients died before their tissues were collected), competing priorities and technical difficulties that inhibit the sharing of information (we only had access to all medical records of one Ogden syndrome patient), as well as establishing the proper controls due to a lack of family members' participation and difficulty in establishing isogenic cells with a base pair correction (Griggs et al., 2009; Ragni et al., 2012). The last issue is even more true if the targeted cell type requires longer time to grow and proliferate such as primary cells especially S37P primary fibroblasts. An alternative approach to overcome the difficulty of establishing the isogenic control fibroblasts can be to correct the mutation in iPSCs and then differentiate these iPSCs into fibroblasts for studying. However, without doubt, our study would be greatly strengthened if the above issues are improved, especially with the addition of more patients and control samples (or use isogenic controls) to each of our assays discussed in the results section (the more lines per group added, the more statistical power will be achieved). I am optimistic that with more and more patients, and their families and physicians recognizing the importance of studying rare disease (especially under the PMI), people will become more enthusiastic about participating in basic research, and more willing to donate tissue or cell samples for mechanism investigation later in the lab.

To follow up on the metabolism alteration including glycolysis elevation seen in the S37P fibroblasts, the following experiments will be helpful to elucidate the mechanism underlying this observation. Use gas chromatography-mass spectrometry to track the metabolites of glucose with C-13 labeling to see whether S37P fibroblasts utilize glucose more rapidly

compared to the control. Use Bioanalyzer to measure the lactate level in the culture media of S37P cells to confirm whether the elevated ECAR is indeed contributed by an accelerated lactate production during glycolysis. Perform the growth curve experiment using decreased concentration of glucose and/or 2-Deoxy-D-Glucose (2-DG, the inhibitor of glucose metabolism) to see whether the growth defect of S37P cells will be further exaggerated (if S37P cells relies more on glucose as main substrate to generate energy). Besides, a parallel biochemical investigation of mitochondrial function-relevant proteins listed in Table 3.2 such as AGC1 would be helpful to find the potential mechanistic player(s). Such experiments might include evaluations of the protein expression (Western blot), protein interaction with NAA10 S37P (IP), protein localization change (IF), and protein stability change in the NAA10 S37P mutant (pulse-chase experiment).

To compensate for the lack of isogenic cells as controls in our experiments, a rescue experiment by introducing the WT NAA10 protein to S37P cells would be useful to confirm whether various defects (or alternations) seen in these cells are indeed caused by the mutant. Furthermore, one could also apply a shRNA or siRNA treatment to the WT cells to demonstrate a general knockdown phenotype of NAA10 or NatA in a different genetic background, which would help show whether the phenotype presented in the S37P mutant is associated with NatA-dependent or independent NAA10 function. For investigating the mechanism underlying specific phenotype seen in S37P cells, one can screen for potential targets by applying strategies including RNAi and targeted proteomics. A very simple example can be to test the hypothesis that RB1 overexpression (discussed in the results) leads to the growth and proliferation defect of S37P fibroblasts. One can knock down RB1 expression (siRNA) in S37P cells and see whether the phenotype is rescued, and/or overexpress RB1 (via a cDNA construct) in the WT cells to see

whether it reduces their growth and proliferation. Of course, this type of experiments can be scaled up to a siRNA library screening if no specific targets are found or hypothesized. For the above example of investigating the growth defect, that means to screen for all proteins in the patient's cells that are known to affect cell growth and proliferation such as cell cycle proteins. Separately, one can also set up experiments such as β-galactosidase staining and telomere length evaluation to investigate whether S37P cells show any premature senescence phenotypes, since these patients do present progeria-like appearance, although lamin staining of the patient's fibroblasts did not reveal any difference compared to the WT control (unpublished data from Max Doerfel).

The evidence we generated in this chapter helped us understand some of the cellular effects of S37P mutation in the patient's fibroblasts. However, since skin (where the fibroblasts were isolated from) is not the major disease-affected organ in Ogden syndrome, the observations we collected in these cells might not reflect what goes on in other tissues and the whole picture of the disease pathogenesis (e.g. the metabolic activity, since skin fibroblasts may exhibit different substrate utilization patterns compared to other tissues that have higher energy demands, such as the heart and brain). Hence, one of the focuses of the future effort will be to establish a robust disease model to investigate major disease-affected tissues in Ogden syndrome, such as the heart and the brain, and to pinpoint the key substrates of NatA or NAA10 that are affected under the S37P background that are responsible for the clinical manifestation of these patients. Modeling the cardiac rhythm disturbances of Ogden syndrome patients will provide us a key to understand the pathogenesis of Ogden syndrome, which I will discuss in detail in the following chapter.

3.5 Concluding Remarks

In conclusion, I reported here a detailed characterization of Ogden syndrome and preliminary studies of fibroblasts from one patient. We provided an in-depth analysis of the clinical phenotype of Ogden syndrome patients, emphasizing on the shared features among all patients. We also introduced the term "NAA10-related disorder", which includes all patients carrying pathogenic NAA10 variants. We demonstrated the highly variable and heterogeneous clinical manifestations of NAA10-related disorder, and we discussed the genotype-phenotype correlation.

The structural modelling data of NAA10 and NatA complex provided evidence of NAA10 S37P affecting the flexibility of NAA10, which may decrease its catalytic properties in both complexed- or free forms. Moreover, the changes in flexibility together with the rearrangement of hydrogen bonds in the interface between NAA10 and NAA15 suggested a reduced NatA complex stability, which was confirmed by our immunoprecipitation experiments where an impaired interaction between NAA10 and NAA15 in S37P cells was clearly seen. Meanwhile, the Nt-acetylome profiling revealed reduction of Nt-acetylation of a subset of NatA substrates due to a reduced catalytic activity, which was hypothesized earlier to be a major contributing factor to the pathogenesis of Ogden syndrome. Our cellular studies suggest altered growth, proliferation and metabolic activities in S37P cells, as well as a mild reduction of NAA10 expression. Ogden syndrome patient's fibroblasts display a possible upregulation of the glycolysis process, which can either be a direct consequence of having a disrupted Nt-acetylation pathway in the cells, or a compensatory change via an indirect effect of the S37P mutation.

The functional data presented here helped to reveal that Ogden syndrome is linked to downstream N-terminal acetylation defects. It provided important insights for future researchers to understand the full mechanism underlying this previously unrecognized syndrome, and contributed to our overall knowledge about the role of Nt-acetylation in development and disease.

Chapter 4

Better understanding cardiac arrhythmia in Ogden syndrome via an iPSC model

4.1 Introduction

Ogden syndrome and long QT syndromes

Cardiovascular disease (CVD) is the leading cause of death for both men and women in the US and worldwide, and it produces immense health and economic burdens. Among all deaths caused by CVD, over 40% are due to sudden cardiac death (SCD), which alone accounts for over

300,000 lives lost per year in the United States (Benjamin et al., 2017). Approximately 5-10% of SCD cases occur in the absence of coronary artery disease or structural heart disease (Zipes et al., 2006), and half of the time the cause is a heritable arrhythmic disorders (Krahn et al., 2009).

The role of NAA10 and protein N-terminal acetylation (NTA) in cardiac arrhythmia pathogenesis, especially in inherited arrhythmogenic syndromes, is literally an untouched field, with only limited evidence (including Ogden syndrome) supporting a possible association of NAA10 mutations and prolonged QT intervals in some individuals of European descents (Casey et al., 2015; Popp et al., 2015; Rope et al., 2011; Saunier et al., 2016). Hence, it is of great importance to establish a disease model of Ogden syndrome to investigate the molecular effect of NAA10 S37P in heart cells, with the hope to help enhance our knowledge about mechanisms of cardiac arrhythmia and heart disease in general, and facilitate the development of new therapeutics to achieve better patient care in the future.

As described above, cardiac dysfunction is one of the most commonly shared features among all NAA10-related disorders especially in Ogden syndrome (OS). OS patients may present both structural and functional defects in the heart (Table 4.1), but it is the cardiac arrhythmia that is usually the life-limiting event. Arrhythmias of OS patients can present in all types, and they are usually primary findings, rather than secondary to infection, effect of medications, or other commonly seen triggers. Among the types of arrhythmias that the OS patients had, prolonged QT intervals, torsade, and T-wave abnormality are extremely striking, since they are rarely seen in normal infants but rather only in those with congenital cardiac channelopathies such as LQTS and catecholaminergic polymorphic ventricular tachycardia (Webster and Berul, 2013). In fact, according to the scoring system for clinical diagnosis of LQTS (Schwartz & Crotti, 2011), Ogden syndrome could be considered as a new type of LQTS,

although the mutation of OS is not in a gene that encodes any of the channel proteins. The genetic screening of OS patients who presented these arrhythmias did not reveal any pathogenic variants in 15 known genes associated with LQTS, although approximately 20% of patients meeting clinical diagnostic criteria for LQTS like OS patients do not have detectable pathogenic variants in one of these 15 genes (Giudicessi and Ackerman, 2013).

Table 4.1 Cardiac features of Ogden syndrome

Utah Family	II-1	II-6	III-4	III-6	III-7
Structural & functional defects	peripheral PS, a septal defect (possibly)	PDA	Small VSD, PFO, trivial PS, trivial bilateral branch pulmonary artery stenosis	ASD/PFO, PDA, mild PHTN, thickened bicuspid aortic valve, dextroposition	PDA, PFO, an enlarged right ventricle, abnormal appearing aortic valve decreased ventricular systolic functions
Arrhythmias & other ECG anomalies	PAC, PVC, SVT, VT	VT, AF	PAC, PVC, TDP, non-specific T wave abnormality, bradycardia	PVC, PAC	VT, bradycardia
Cardiomegaly	yes	yes	Mild (left atrium)	yes	Yes (right side is bigger)
Age of death (m)	11 1/2	9 1/2	15	4 1/2	5 1/2

AF, atrial fibrillation; ASD, atrial septal defect; PAC, premature atrial contraction; PDA, patent ductus arteriosus; PFO, patent foramen ovale; PHTN, pulmonary hypertension, newborn; PS, pulmonary stenosis; PVC, premature ventricular contraction; SVT, supraventricular tachycardia; TDP, Torsades de pointes; VSD, ventricular septal defect; VT, ventricular tachycardia.

Prevalence of LQTS is estimated at 1 per 2000 live births (Schwartz et al., 2009). Most types of LQTS do not involve phenotypes beyond characteristic prolonged QT intervals on 12-lead electrocardiograms (ECGs), with only a few exceptions including long QT syndrome type 8 (LQT 8 or Timothy syndrome), which is the most severe phenotypic form of LQTS. Like OS, LQT 8 is quite rare, and patients manifest a complex phenotype and usually have a very short life span (2.5 years on average). In fact, patients of OS and LQT 8 share many clinical features, including dysmorphic facial features, developmental delay, hypoglycemia, congenital heart defects (PDA, PFO, VSD, pulmonary artery stenosis and cardiomegaly), arrhythmias (PVCs, QT prolongation, VT, and bradycardia), hypotonia, recurrent infections (ear, respiratory, post-op, etc.) and sudden death. The genetic basis of LQT 8 is gain-of-function mutations in gene CACNA1C, which encodes the pore-forming $\alpha 1$ subunit of $Ca_V1.2$ channel, the main L-type channel expressed in the mammalian heart, and also widely distributed in many other tissues (Liao and Soong, 2010). The close resemblance of the clinical manifestations between the two disorders might hint at an overlapping dysregulated molecular pathway (Gelb & Tartaglia, 2006; Oti & Brunner, 2006).

In fact, as predicted from N-terminus amino acid sequences, many channel proteins in which mutations have been discovered to cause or associate with cardiac channelopathies are targets of NatA modification (see an incomplete list in Table 4.2). Notably, mutations found in half of proteins listed here can cause prolonged QT intervals, i.e. cardiac ryanodine receptor (RyR2) (NH_2-"$MA_{DGGEGEDE}$"), a very important player in regulating calcium flux in the heart and contributing to multiple types of arrhythmias including LQTS (Medeiros-Domingo et al., 2009); and calmodulin (CaM) (NH_2-"$MA_{DQLTEEQI}$"), which interacts with RyR2 and has been proven to contribute to LQTS (Limpitikul et al., 2014; Søndergaard et al., 2015). It is likely that

one or multiple ion channel proteins and their key regulators might be affected by NAA10 S37P and exert dysregulated functions during embryonic development and postnatal performance of the cardiovascular system, thus giving rise to the complex cardiac phenotype of OS patients and finally triggering their deaths during infancy.

Table 4.2 Predicted targets of NatA linked to channelopathy diseases

Channelopathy	Gene	Protein	Functional role in cardiomyocytes
ARVC 10	*DSG2*	DSG2	desmosomal protein
ARVC 5	*TMEM43*	TMEM43	transmembrane protein
ARVC 8	*DSP*	DSP	desmosomal protein
ARVC 9	*PKP2*	PKP2	desmosomal protein
BrS 2	*GPD1L*	G3PD1L	not fully established
BrS 4	*CACNB2*	$Ca_v\beta2b$	β2b subunit of $I_{Ca,L}$ channel
BrS 5	*SCN1B*	$Na_v\beta1$	β subunit of I_{Na} channel
CPVT 1/ARVC 2	*RYR2*	RyR2	Ryanodine receptor in sarcoplasmic reticulum membrane
LQT 1/SQT 2	*KCNQ1 (KVLQT1)*	$K_v7.1$ (K_vLQT1)	α subunit of I_{Ks} channel
LQT 12	*SNTA1*	α1-syntrophin	scaffolding protein
LQT 13	*KCNJ5*	$K_{ir}3.4$	α subunit of $I_{K,ACh}$ channel
LQT 3/BrS 1/PCCD	*SCN5A*	Na_v 1.5	α subunit of I_{Na} channel
LQT 4	*ANK2*	Ankyrin-B	Anchoring protein
LQT 6	*KCNE2*	KCNE2 (MiRP1)	β subunit of I_{Kr} channel
LQT 7/SQT 3	*KCNJ2*	K_{ir} 2.1	α subunit of I_{KI} channel
LQT 8/SQT 4/BrS 3	*CACNA1C*	$Ca_v1.2$	α1C subunit of $I_{Ca,L}$ channel
BrS*	*KCND3*	$K_v4.3$	α subunit of I_{to} channel
LQT*	*TRDN*	Triadin	links calsequestrin-2 and RyR2
SQT 5	*CACNB2b*	$Ca_v\beta2b$	β2b subunit of $I_{Ca,L}$ channel
SQT 6	*CACNA2D1*	$Ca_v\alpha2\delta$-1	α2δ subunit of $I_{Ca,L}$ channel
WPW	*PRKAG2*	PRKAG2	protein kinase
CPVT 4/LQT 15	*CALM1*, *CALM2*	Calmodulin (CaM)	intracellular Ca^{2+} sensor

LQT, long QT syndrome; SQT, short QT syndrome; CPVT, catecholaminergic polymorphic ventricular tachycardia; ARVC, arrhythmogenic right ventricular cardiomyopathy; WPW, Wolff-Parkinson-White; PCCD, progressive cardiac conduction disease; BrS, Brugada syndrome; $I_{Ca,L}$: L-type calcium current; $I_{K,ACh}$: acetylcholine-sensitive potassium current; I_{KI}: inward rectifier potassium current; I_{Kr}: rapid delayed rectifier potassium current; I_{Na}: fast sodium current. *, association discovered (Giudicessi et al., 2011).

iPS-CM disease modeling

One of the great challenges to investigate the effect of disease-associated variants *in vitro* is the lack of disease-affected tissues or cell samples, especially from the heart, which can only be obtained through invasive procedures or post-mortem. Mouse models are widely used for disease study and have many merits; however, they are far from ideal for modeling human arrhythmogenic disorders due to many electrophysiological differences between the two species (Bader et al., 2011; Cheng et al., 2011; Splawski et al., 2004; Thiel et al., 2008). The discovery of human induced pluripotent stem (iPS) cells has brought new hope to this field, since it enables researchers to generate an individualized patient iPS cell-derived "disease-in-a-dish" (Gage, 2010).

iPS technology was first developed in 2006. By simultaneously introducing four transcription factors (Oct3/4, Sox2, Klf4 and c-Myc) via viral transduction, researchers were able to generate cells with properties similar to embryonic stem cells (ESCs) from mouse fibroblasts, and these were later named induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka, 2006). The next year, human iPSCs were generated from human fibroblasts using a similar approach (Takahashi et al., 2007). Since then, many laboratories have joined the effort to explore the underlying mechanisms of somatic cell reprogramming and gradually optimize the iPSCs generation procedures. To date, many methods have been developed to generate iPSCs from various types of somatic cells, among which the integration-free strategies, such as mRNA transfection or Sendai virus (SeV) transduction have drawn more and more attention due to the lack of insertional mutagenesis and oncogenic transformation risks (Bayart and Cohen-Haguenauer, 2013).

Human (i)PSCs display similar characteristics to human embryonic stem cells (hESCs); they can thus give rise to cell lineages of all three germ layers including cardiomyocytes (Zhang et al., 2009; Zwi et al., 2009). Over the past decade, various methods have been developed to enable the differentiation of hPSCs into functional cardiomyocytes, which can be generalized into two categories, embryoid bodies (EBs)-based (Graichen et al., 2008; Yang et al., 2008; Zhang et al., 2009) and monolayer-based differentiation (Laflamme et al., 2007; Zhang et al., 2012). These methods either rely on serum-based spontaneous differentiation or the addition of chemical factors or small molecules to direct the differentiation to cardiac lineages (Bellin et al., 2012; Burridge et al., 2014; Burridge et al., 2012). Generally speaking, monolayer-based differentiation protocols have many merits over EBs-based ones for their relatively lower cost, easier manipulation, and higher efficiency in the year of 2013, when we started this project (reviewed in Mummery et al., 2012).

In fact, many of inherited arrhythmogenic syndromes and cardiac channelopathies have been successfully modeled by iPSCs (reviewed in Bezzerides et al., 2017; Goedel et al., 2017), including long QT syndrome (LQTS) (Itzhaki et al., 2011; Malan et al., 2011; Moretti et al., 2010; Yazawa et al., 2011; Yazawa & Dolmetsch, 2013), catecholaminergic polymorphic ventricular tachycardia (CPVT) (Fatima et al., 2011; Jung et al., 2012; Novak et al., 2012), Brugada syndrome (Davis et al., 2012; Okata et al., 2016), and arrhythmogenic cardiomyopathy (Caspi et al., 2013; Kim et al., 2013; Ma et al., 2013). The knowledge we gained from these studies provides great value not only for deciphering disease biology, but also in the field of safety pharmacology.

The work discussed in this chapter aimed to create a disease model of Ogden syndrome using patient specific iPS cell-derived cardiomyocytes (iPS-CMs) to investigate the cardiac

phenotype of Ogden syndrome at a cellular level, and shed light on the mechanism of Ogden syndrome pathogenesis.

4.2 Materials and Methods

Ethics and consent

The cell line creation and manipulation were under protocols approved by the Institutional Review Board at Cold Spring Harbor Laboratory (CSHL IRB-012-009). Written informed consent was obtained by others from individuals and families of study for sample collections and publication. All medical records were de-identified before reviewing. All experiments were performed in accordance with the approved guidelines and regulations.

Matrigel plate coating

Matrigel matrix (BD Biosciences, # 354230) was thawed according to manufacture instruction and then diluted using cold DMEM/F12 (Thermo Fisher Scientific, # 11330057) on ice to a final concentration of 100 μg/ml. 1× Matrigel solution was plated on tissue culture-treated plates or dishes and ensured the coverage of the entire bottom of the well or dish. The coated plates are placed in a biological safety cabinet at room temperature for 1 h with the lid on (to prevent dehydration). Then, Matrigel solution is aspirated, and iPSCs are plated immediately.

Feeder-free iPSC derivation and cell culture

Primary fibroblasts WT-1 control and S37P Ogden syndrome were obtained and cultured as described in 3.2 method section above. WT-4 and WT-6 hiPSC lines were purchased from Harvard Stem Cell Institute (#BJ-SiPS-C and #BJ-SiPS-D). Fibroblasts prepared for

reprogramming experiments were first tested for mycoplasma using e-Myco plus Mycoplasma PCR Detection Kit (iNtRON Biotechnology, # 25237) according to the manufacturer's protocol. Fluorescence in situ hybridization (FISH) analysis and karyotyping were carried out by WiCell Cytogenetic Testing Services and Mount Sinai Genetic Testing Laboratory to confirm the gender of the fibroblasts as well as to rule out general chromosomal abnormalities.

For iPSC induction, Sendai virus-based reprogramming was carried out using P4 mycoplasma-free fibroblasts on 6-well plates according to the manufacturer's protocol (CytoTune™-iPS 2.0 Sendai Reprogramming Kit, Thermo Fisher Scientific, # A16517) in collaboration with Sunita D'Souza at Mt. Sinai. Briefly, 1 million fibroblasts were transduced at an optimized MOI (5, 5, and 3 for KOS, hc-Myc, hKlf4) of SeV 2.0 vectors in DMEM medium containing 20% fetal bovine serum (FBS). And on day 8, 400K cells were re-plated on 10 cm cell culture dishes covered with irradiated mouse embryonic fibroblasts (MEFs, feeder) cells. In the next four days, cell culture medium was gradually switched to hESC medium (DMEM/F12 supplemented with 20% Knockout Serum Replacement (Thermo Fisher Scientific, # 10828028), 2 mM L-glutamine (Thermo Fisher Scientific, # 25030081), 0.1 mM Non-Essential Amino Acids (Thermo Fisher Scientific, # 11140050), 55 μM β-mercaptoethanol (Thermo Fisher Scientific, # 21985023), and 10 ng/mL FGF-Basic (AA 10-155) Recombinant Human Protein (Thermo Fisher Scientific, # PHG0026)). Several colonies with typical iPSC morphology and positive staining for TRA-1-60 (Stemgent, # 09-0068) were picked around 4 weeks of induction and transferred to 6-well plates with the presence of 2 μM ROCK inhibitors (Y27632 (Stemgent, # 04-0012) or Thiazovivin (EMD Millipore, #420220)), which was removed within 24 hr. iPSC colonies were then manually propagated until transferring to a feeder-free culture condition around passage six.

Feeder-free iPSC culture was maintained on 6-well plates coated with Matrigel and cultured in mTeSR™1 medium (Stemcell technologies, # 05850), supplemented with mTeSR™1 5X Supplement (Stemcell technologies, # 05852) and 2 mM L-glutamine (Thermo Fisher Scientific, # 25030081). Routine passage of iPSCs under feeder-free condition was achieved by dissociating with 0.5 mM EDTA (Thermo Fisher Scientific, # 15575020) and re-plating directly without centrifuging to a freshly Matrigel-coated well of a 6-well plate. All cells were cultured in an incubator at 37 °C, 5% CO2.

iPSC characterization

Two months into passaging iPSC colonies, cells were confirmed for the absence of Sendai virus (SeV) (by Pluripotent Stem Cells Shared Resource Facility at Mt. Sinai) via an immunocytochemistry study with anti-SeV antibody (MBL, # PD029) as well as by qPCR targeting the SeV genome and transgenes. SeV-free iPSCs then underwent chromosomal G-banding analysis (karyotyping) using a standard procedure (by WiCell Cytogenetic Testing Services and Mount Sinai Genetic Testing Laboratory) to rule out chromosome abnormalities. iPSC lines with normal karyotypes were further assessed for pluripotency by immunostaining and real-time quantitative PCR (qPCR) for endogenous pluripotency markers, and by an *in vitro* differentiation (embryoid body) assay. The EBs generated from the WT iPSCs were evaluated by histology.

For qPCR for detecting pluripotent markers, total RNA of iPSCs was extracted using PureLink® RNA Mini Kit (Thermo Fisher Scientific, # 12183018A) and then cDNA was produced using a SuperScript® III First-Strand Synthesis System (Thermo Fisher Scientific, # 18080051) and analyzed using Fast SYBR Green Master Mix (Thermo Fisher Scientific, #

4385612) according to manufacturer's instruction in a LightCycler 480 (Roche). The comparative Ct (ΔΔCt) Method was used to analyze the relative change of gene expression in iPSC lines compared to the in-house human embryonic stem cell lines H1 and H9 using β-actin as a reference gene. The primer sets used in qPCR are as follows (5' - 3', forward/reserve):

OCT4: aacctggagtttgtgccagggttt/tgaacttcaccttccctccaacca
NANOG: cctgaagacgtgtgaagatgag/gctgattaggctccaaccatac
SOX2: agaagaggagagagaaagaaagggagaga/gagagaggcaaactggaatcaggatcaaa
KLF1: tgtttggtggtctcttcacacgga/gggtttgcacgacagtttggacat
TERT: tgaaagccaagaacgcagggatg/tgtcgagtcagcttgagcaggaatg
β-ACTIN: tttttggcttgactcaggattt/gcaagggacttcctgtaacaac

In vitro differentiation of iPSCs to three germ layers

WT-1 #2 and S37P-H iPSCs with passage number 25-27 were harvested from Matrigel-coated 6-well plates by dissociating monolayer cultures using Accutase (Innovative Cell Technologies, # AT104). Then Accutase was aspirated, and cells were collected with DMEM/F12 (GE Healthcare Life Sciences, # SH30023.FS) in 15 ml conical tubes. 200× Matrigel was added to cell suspension, and the cells were incubated in it at RT for 15 min. After centrifugation at 200 g (1,000 rpm) for 5 min, the cell pellet was incubated in the tube for another 15 min. The supernatant was carefully aspirated, and the cell pellet was gently re-suspended using EBM medium (DMEM/F12 supplemented with 20% Knockout Serum Replacement, 0.1 mM non-essential amino acids, 2 mM L-glutamine, 55 μM 2-mercaptoethanol) with 2 μM thiazovivin (EMD Millipore, # 420220). Then, the cell clumps were transferred to wells of an ultra-low attachment plate (Corning # 3471). After 24 hours of incubation, the

resultant EBs were collected and changed to fresh EBM medium by gravity-induced sedimentation in 15 mL conical tubes. Suspension cultures were maintained by replacing fresh EBM medium in the same fashion every other day for 5 weeks.

Histological analysis of embryoid bodies

Embryoid bodies (EBs) were pooled at week 5 of differentiation, rinsed in PBS and fixed in 4% PFA at 4℃ overnight. Fixed EBs were rinsed 3 times in PBS and embedded in Histogel® (Thermo Fisher Scientific, #HG-4000-012, given by Dr. Rasmani Hazra at CSHL) at RT. Samples were processed, embedded, sectioned and hematoxylin and eosin (H & E) stained by the histology shared resource at CSHL. H & E stained sections were imaged using Aperio Light Field Slide Scanner microscope (Leica Biosystems).

Phase contrast image analysis

To examine morphological features of iPSCs and iPSC-derived cardiomyocytes, phase contrast images of iPSCs and iPSC-derived cardiomyocytes adherent cultures were obtained and analyzed using an inverted phase contrast microscope (Zeiss). ImageJ software was utilized for image analysis.

Immunofluorescence

iPSCs were fixed with 4% PFA for 15 minutes at RT, followed by two rinses with PBS. Next, cells were permeabilized with 0.2% Triton-X-100 in Saponin solution (1× PBS, 5 mg Glucose, 0.2% BSA, 0.2% NaN_3, 5 mM HEPES, 0.5% Saponin, and added Milli-Q water to go up to the desired volume) for 5-minute at RT, followed by two rinses in PBS with gentle rocking.

Then, cells were incubated in minimum volume of the blocking buffer (1× PBS, 10% normal goat serum (Cell Signaling, # 5425), 0.1% Saponin) for 60 minutes at RT. Cells were then incubated with indicated primary antibody in 1× PBS/1% BSA/0.1% Saponin overnight at 4°C. The cells were rinsed three times (10-minute each) with above solution at RT with gentle rocking. Then, cells were incubated in dark with fluorescent conjugated secondary antibody in 1× PBS/1% BSA/0.1% Saponin for 60 minutes at RT. Then secondary antibody was aspirated, followed by the same nuclear stain, mounting and imaging study as described in section 3.2 under IF.

For iPS-CMs, cell membrane permealization was achieved by incubating samples in 0.2% Triton X-100 in 5% FBS for 5 min. 1% FBS was used for blocking, antibody dilution and washing solution. After antibody staining, cell nuclei were stained with 1 μg/ml DAPI with 10 min incubation in PBS. Imaging was collected using the Olympus FluoView FV1000 confocal system. Gain was kept constant for WT control and patient groups to normalize and exclude auto-fluorescence contributions.

Antibodies

Primary antibodies include anti-hOct4 (BioVision, # 3576-100, 20 μg/ml), anti-hNANOG (R&D Systems, # AF1997, 10 μg/ml), anti-hSox2 (Thermo Fisher Scientific, # 48-1400, 2 μg/ml), TRA-1-60 (Stemgent, # 09-0068, 1:100), TRA-1-81 (Stemgent, # 09-0011, 1:100), anti-SSEA-4-Alexa Fluor® 647 (BD Biosciences, # 560219), anti-α-actinin (Sigma-Aldrich, # A-7811, 1:600), anti-vimentin (Abcam, # ab137346, 1:800), and anti-connexin 43 (EMD Millipore, # AB1728, 1:300), rabbit anti-GAPDH (Santa Cruz, # sc-25778, WB 1:500), rabbit anti-NAA10 (Protein Tech, # 14803-1-AP, 1:1000), mouse anti-NAA15 (abcam, # ab60065, 1:100). Secondary antibodies include AlexaFluor488 goat anti-mouse (Thermo Fisher Scientific, # A-

11029), Alexa Fluor 488 donkey anti-rabbit (Thermo Fisher Scientific, # A-21206, 1:100), AlexaFluor594 goat anti-rabbit (Thermo Fisher Scientific, # A-11037), Alexa Fluor 647 donkey anti-goat (Thermo Fisher Scientific, # A-21447, 1:1000), and Alexa Fluor 647 goat anti-mouse (Invitrogen, # A21235, 1:1000) for immunofluorescences. Anti-mouse IRDyse 800CW (LI-COR, # 926-32210, 1:15000) and anti-rabbit IRDye 680RD (LI-COR, # 926-68071, 1:15000) were used for western blotting. For Western blotting method please see 3.2 Materials and Methods section.

In vitro cardiomyocyte differentiation and culture

I learned how to perform this assay from Dr. Sandra Klein in Dr. Christoph Schaniel's lab at Mt. Sinai. All results discussed in this dissertation were based on cardiomyocytes that I independently generated. Before differentiation, all target iPSCs were maintained in feeder-free condition with high qualities for pluripotency. *In vitro* cardiomyocyte differentiation was performed using a monolayer-based method with adaptation based on the inhibition of GSK3 and Wnt pathways (Lian et al., 2012). Briefly, iPSCs were collected using EDTA method and seeded on Matrigel-coated plastic bottomed 12-well plates at about 20% confluency. For high throughput characterization using OptoDyCE platform, iPSCs were seeded directly on Matrigel-coated glass bottomed 96-well plates. To gain the best differentiation outcome, iPSCs were seeded no longer than 4 days prior to the start day of the differentiation (Day 0), with an even distribution of clones in each well.

When the cells reached over 90% confluency, the differentiation was initiated by adding CHIR99021 (Stemgent, # 04-0004) to an optimized concentration (8 μM) in differentiation medium (RPMI 1640 (Thermo Fisher Scientific, #11875-093) supplemented with B27 without

insulin (Thermo Fisher Scientific, # A1895601)). Cells were cultured in a 37°C, 5% CO2, and 5% O2 incubator. CHIR was removed 24 hours later, and cells were cultured in fresh differentiation medium for another 48 hours. On day 3 of differentiation, IWP2 (Stemgent, # 04-0034) was added to the cells at a final concentration to 5 μM in combined medium (1 volume of the old differentiation medium from each well mixed with equal volume of the fresh differentiation medium) for 48 hours. At day 5, the medium with IWP2 was replaced with fresh differentiation medium. At day 7, medium was switched to maintenance medium (RPMI 1640 supplemented with complete B27 (Thermo Fisher Scientific, #17504044)). Cells were cultured in the 37°C, 5%CO2, and 5% O2 incubator until day 12 when transferred to a normal oxygen incubator (37°C, 5% CO2/air). Medium was replaced every 3 days after day 7. Robust spontaneous contraction normally occurred by day 12 (could be seen as early as day 7 when differentiated in 96-well plates). The cells can be maintained with this spontaneously beating phenotype for more than 6 months.

Re-plating iPS-derived cardiomyocytes

To prepare iPS-CMs for downstream assays including electrophysiology and immunofluorescence studies, iPS-CMs were microdissected and re-plated on glass bottomed plates or dishes using the following method. Dishes meant to receive the dissociated cells were coated with $PBS^{Ca/Mg}$ + fibronectin (10 μg/mL, Sigma, # F1141) overnight at 4°C or 37°C for 3 hours. The explants were scratched off using the tip of a 10 mL syringe and cut into small pieces. Then, the explants were transferred into a 1.5 mL Eppendorf tube with iPS-CMs maintenance medium. The explants settled to the bottom of the Eppendorf and then medium was removed. The explants were washed twice with 500 μM HBSS (Thermo Fisher Scientific, # 14170), and

then once with Collagenase solution (1.5 mg/mL Collagenase Type 2 (Worthington Biochemical, # LS004174) in HBSS). The explants were incubated with Collagenase solution in a pre-warmed Thermomixer for 30 min (37°C, 850 rpm). After half an hour, the dissociated cell supernatant was transferred into a prepared centrifuge tube with iPS-CMs differentiation medium. Fresh Collagenase solution was added to the original Eppendorf containing the non-dissociated explants. The process was repeated until the explants fully dissociated. The dissociated cells were spinned down for 5 min at 250-300 g. The supernatant was carefully removed. The pellet was re-suspended in iPS-CMs maintenance medium, and cells were plated on fibronectin-coated dishes. Fresh media was changed on day 2 after dissociation. Assays were carried out within a week after the cells were re-plated.

Electrophysiology studies of iPS-CMs

These experiment were done in collaboration with Dr. Entcheva's lab. In particular, Aleks Klimas handled the OptoDyCE-related apparatus for recording and imaging experiments and helped with the addition of virus to iPS-CMs. iPS-CMs of 1, 5, and 6 months old were optogenetically modified to express ChR2 (H134R) via an adenoviral delivery method (MOI: ~250) as previously described (Ambrosi & Entcheva, 2014; Williams et al., 2013). 48 hours after the infection, membrane voltage (Vm) and intracellular calcium ($[Ca2^{+}]_i$) of the infected cells were optically recorded using the synthetic voltage-sensitive dye Di-4-ANBDQBS (Di-4, from Dr. Leslie Loew, University of Connecticut) and calcium indicator Quest Rhod4 AM (Rhod4, AAT Bioquest, # 21120), respectively, with OptoDyCE following a previously described protocol (Klimas et al., 2016).

Briefly, for Vm recording, a 17.5 mM Di-4 stock solution in pure ethanol was diluted to 35 μM fresh Tyrode's solution (135 mM NaCl, 4 mM KCl, 1 mM $CaCl_2$, 2 mM $MgCl_2$, 5 mM glucose and 5 mM HEPES, pH 7.4). Cells were stained for 6 min in dye solution followed by a 6 min wash in fresh Tyrode's. This wash solution was then replaced by fresh Tyrode's. And for $[Ca2^{+}]_i$, a 0.5 mM Rhod4 stock solution dissolved in DMSO with 20% Pluronic to 10 μM in Tyrode's solution. Samples were stained with this solution for 20 min, followed by a 20 min wash in fresh Tyrode's solution, and finally a replacement with fresh Tyrode's before experiments.

Optical imaging was performed at > 200 frames per second (fps) with 4×4 binning using iXon Ultra 897 EMCCD camera (Andor Technology). Optical stimulation (470 nm) was provided at pulse lengths of 10 ms, at 0.5-4 Hz, with irradiances of 0.624-3.457 mW/mm^2, as needed. Electrical stimulation was performed using 3-month old iPS-CMs on 14-mm glass-bottom dishes, delivered through a pair of parallel platinum electrodes connected to a pulse generator (IonOptix, Milton, MA) providing 5 ms 10V bipolar pulses at 0.25-1.5 Hz.

iPS-CMs of 1 month old also underwent mexiletine (50 μM) test performed by Aleks Klimas at Washington D.C. Upon finishing recording, all infected cells were fixed and subject to immunofluorescence analysis using the Olympus FluoView FV1000 confocal system.

iPSC lines used in each characterization assay

iPSC lines	Genotyping	qPCR	Karyotyping	IF	EB formation	iPS-CMs studies
WT1-#1	+	+	+ abnormal	+	-	-
WT1-#2	+	+	+	+	+ (histology)	+
WT1-#5	+	+	+ abnormal	-	-	-
(WT) BJ-SiPS-C	+*	+*	+*	+*	+* (qPCR)	+*
(WT) BJ-SiPS-D	+*	+*	+*	+*	+* (qPCR)	+*
S37P-D	+	+	+	-	-	+
S37P-E	+*	+*	+*	-	-	-
S37P-H	+	+	+	-	+* (qPCR)	+
S37P-I	+	+	+	+	-	+
S37P-L	+*	+*	+* contains a normal variant	-	-	-
S37P-K	+*	+*	-	+*	-	-

BJ-SiPS-C and D lines were characterized by HSCI, but iPS-CMs were derived from them by me. *, data was not included or discussed in this dissertation. IF, immunofluorescence; EB, embryoid body.

Data process and statistical analysis

Electrophysiology data collected from OptoDyCE platform was processed and analyzed using flashligHTS (which has not been commercialized) by Aleks Klimas as described previously (Klimas et al., 2016).

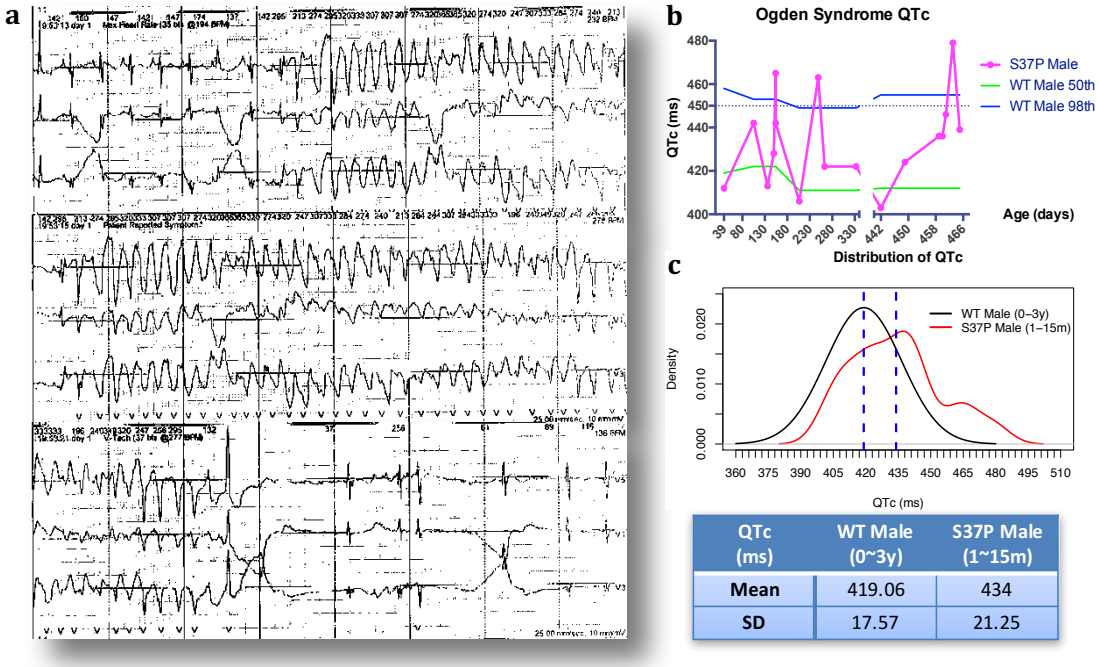
For both action potential and calcium transient recordings, cell clones of each group (patient vs WT control) were split into multiple dishes as replicates, multiple field of views (FOVs) were taken per dish during recording, and each recording was divided into smaller region of interest (ROI) and intensity data over time was extracted and analyzed. Contractions were measured by post-processing of recorded videos and tracking cell motion (displacement) by naturally occurring cell heterogeneities (fiducial markers). All experiments were repeated multiple times on different dates to ensure the reproducibility. Electrophysiology data are presented in terms of mean ± standard error of the mean. Statistical analysis was done using two-way ANOVA with Tukey-Kramer post-hoc test for the comparison between patient and WT control groups by Aleks.

4.3 Results

Ogden syndrome patient (Utah, III-4) has high QT variability

The electrocardiographic QT interval is from the onset of the QRS complex to the end of the T wave, which is the interplay of several ion channels and is a reflection of the summed action potential durations in the heart ventricles. A prolonged (corrected) QT interval (QTc) and a high QTc variability commonly seen in LQTS patients have been associated with life-threatening ventricular tachyarrhythmias such as Torsades de Pointes (TdP), sudden cardiac death and overall mortality (Goldenberg et al., 2006; Moss et al., 1991; Mozos & Caraba, 2015). Analysis of a 24 h Holter monitor record of an Ogden syndrome patient (Utah family III-4) indeed documented a non-sustained episode of Torsade de pointes (polymorphic ventricular tachycardia) (Fig 4.1 a), and the symptoms reported at the time included screaming, turning colors, and holding breath. In order to characterize the QT intervals of Ogden syndrome patients, we collected all 12-lead ECG traces available from this patient. In total, 17 ECG traces between 1-month old and 15-month old of age were obtained and analyzed. From these recorded ECG traces, the duration of the QT interval was assessed from lead II and corrected for heart rate using Bazett's formula (Bazett, 1920).

$$QTcB = \frac{QT}{\sqrt{RR}}$$



QTc (ms)	WT Male (0~3y)	S37P Male (1~15m)
Mean	419.06	434
SD	17.57	21.25

Figure 4.1 ECG analysis of an Ogden syndrome patient. (a) 24 h Holter monitor of Utah III-4 patient at 14.5 months old documented a self-terminating episode of Torsade de pointes. (b-c) Further ECG analysis of this patient revealed a high QTc variability during a serial of ECG follow-ups from one to 15 months old. For the purpose of comparison, ECG data from the same gender and age group was pulled (Rijnbeek et al., 2001) and illustrated in the same graph as Ogden syndrome patient. ECG traces shown in (a) were extracted from one Ogden syndrome patient's (Utah family, III-6) Holter monitor recording, which data was collected by Dr. Lyon.

Within this ECG data collection, the minimum QTc value is 403 ms at 14 months old of age, the maximum value is 479 ms at 15 months old, and the median is 436 ms. Notably, even within a short period of time, patient's QTc values fluctuated quite dramatically, such as a difference of 76 ms detected within 21 days, and 37 ms seen in four days. And even within the same day, a QTc difference of 23 ms was captured on two separate ECGs (Fig 4.1 b). In order to visualize and illustrate the degree of high QTc viability of this patient, I also analyzed the ECG data from the same gender and age group for comparison (Rijnbeek et al., 2001). The distribution of the patient's QTc not only has a higher mean (434 vs 419 ms) but also a bigger standard deviation (21.25 vs 17.57 ms) compared to the population data (Fig 4.1 c). Usually one would expect the variation observed within a population to be bigger than that within an individual. Hence, this analysis provides additional evidence to support that Ogden syndrome patients tend to have a high QT variability, and in other words, low cardiac electrical stability.

NAA10 substrates are enriched during reprogramming

When I first tried to create an iPSC model for Ogden syndrome using mRNA-based reprogramming method (mRNA Reprogramming Kit, Stemgent, #00-0071), I experienced many challenges (of reprogramming both WT and S37P fibroblasts) along the way including very low transfection efficiency (Fig 4.2 a), cell death (seen more frequently in the patient cells) when exposed to the reagents, and extremely low reprogramming efficiency. As a matter of fact, we only got about 3~5 successful iPSC-like colonies showing up from around 200K WT-0 control fibroblasts, which eventually merged to a big colony due to the proximity of these small colonies to each other (Fig 4.2 b). And no matter how hard I optimized the RNA-mediated reprogramming experiments (four rounds were tried) as well as changing to Sendai virus based reprogramming method (tried one round with the CytoTune™-iPS 1.0 Sendai system), we were

not successful in boosting the reprogramming efficiency and could not obtain more iPSC-like colonies from either our WT or S37P fibroblasts. We did, however, successfully obtain iPSC clones during the same experiment from another WT control fibroblasts (HDFn) that we imported from our collaborator (data not included). A major difference between HDFn and our fibroblasts (WT-0) was the passage number; our donor cells were at P12-P15 in culture (with patient fibroblasts being 3 passage older than the WT), whereas the HDFn fibroblasts were only at P2. It is evident now that increased donor age and late passage number in culture are major negative factors for reprogramming efficiency, due to the upregulation of the cell cycle inhibitory protein p21 induced by the associated cellular senescence (Trokovic et al., 2015).

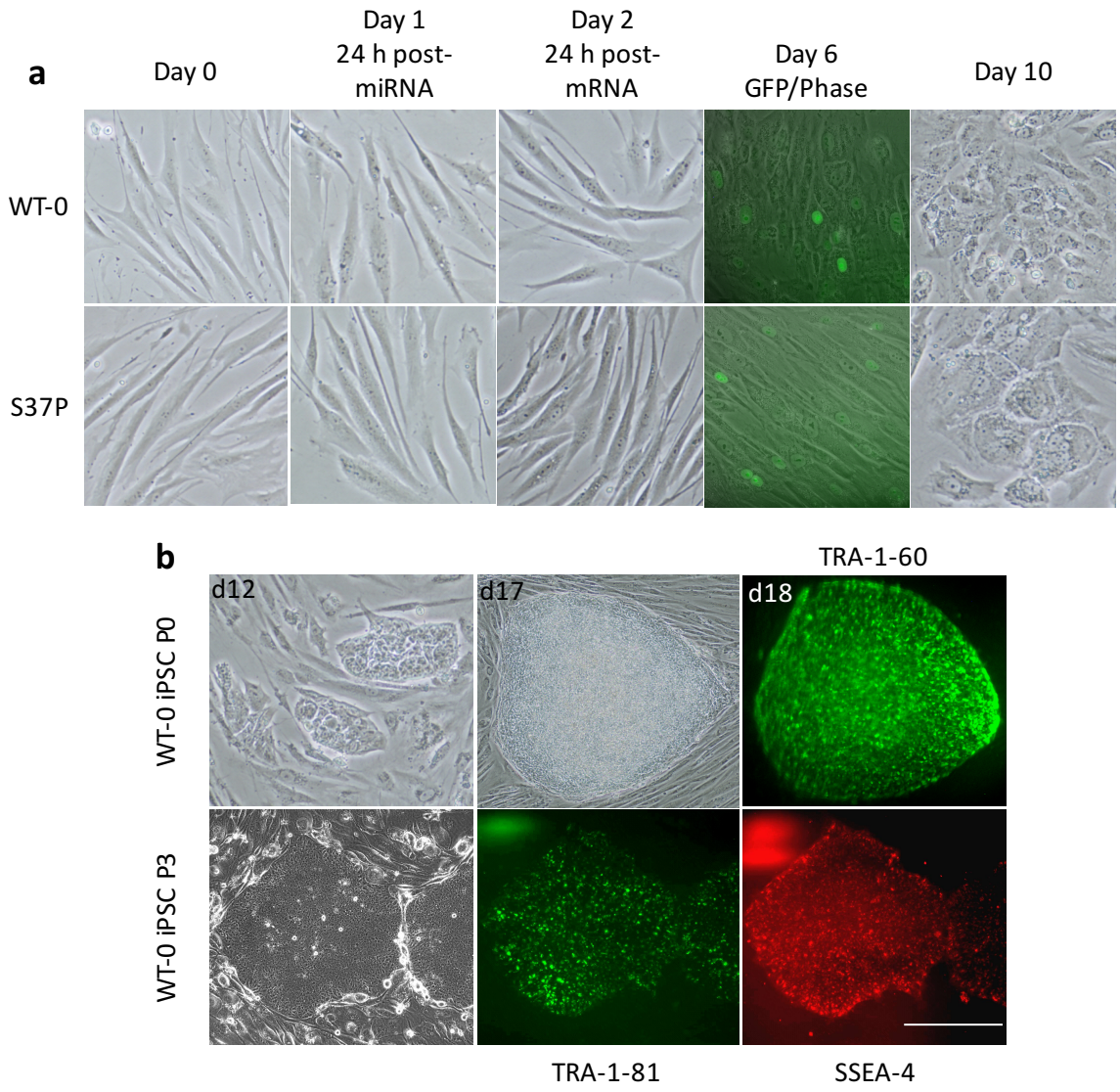


Figure 4.2 Generation of RNA-induced pluripotent stem cells from primary fibroblasts. (a) Morphology change of primary fibroblasts between day 0 to day 10 of reprogramming was monitored by bright field microscopy (images: 10×). (b) Bright-field images taken during the derivation of RNA-iPSCs from WT-0 fibroblasts showing small hESC-like colonies (day 12) and appearance of a mature iPSC clone (day 17). Immunocytochemical analysis of pluripotency marker TRA-1-60 on the same WT-0 iPSC clone (day 18) and two additional pluripotency markers in passage 3 cells after mechanical picking and expansion. Scale bar represents 100 μm. miRNA, microRNA. RNA-induced reprogramming experiments were repeated four times.

While troubleshooting the reprogramming experiments, one of my concerns included whether S37P fibroblasts would have any defect during reprogramming *in vitro* due to dysregulation of key factors required during such a dynamic and complex process (Apostolou & Hochedlinger, 2013; Hochedlinger & Jaenisch, 2015; Li & Belmonte, 2017). In order to get some insights into this question, I designed and carried out a proteome-wide analysis to profile NatA substrates during reprogramming of somatic cells to pluripotency. (Bioinformatics assistance for extracting all MS-, MA-, MT- starting proteins and their sequences from Uniprot database was obtained from Laura Jimenez.)

Due to the lack of proteomics data during human cell reprogramming at the time, we analyzed the mouse proteome instead for an estimation. The entire annotated mouse proteome consisting of 74,181 proteins was downloaded from UniProt release 2013_12 (http://www.uniprot.org/), and the first 10 amino acid sequences at the N terminal were extracted for analysis. Among them, 7,918 unique protein groups that expressed were confidently identified during reprogramming of mouse embryonic fibroblasts (MEF) to iPSCs by high-resolution nano liquid chromatography-tandem mass spectrometry (LC-MS/MS) (Hansson et al., 2012). Three major NatA representative substrate groups, proteins starting with "MA" (MA-), "MS" (MS-), or "MT" (MT-) (Helbig et al., 2010), were selected from the total mouse proteome list and the reprogramming list for comparison.

As seen, MA-, MS- and MT- N termini proteins account for about one third of the mouse proteome in total. And a significant enrichment of MA- or MS- N termini proteins during reprogramming was seen (Fig 4.3 a-c), which in combination account for 41.43% of all proteins detected during reprogramming. MA- and MS- N termini proteins comprise the majority (about 70%) of NatA substrates, and among them, 95% of MA- and 99% of MS- starting proteins are

acetylated by NatA *in vivo* (Aksnes et al., 2016). Considering the proteome similarity between human and mouse and the evolutionary conservation of N-terminal acetylation in eukaryotic cells (Aksnes et al., 2016; Arnesen et al., 2009), we estimate that about 3157 out of 7918 proteins (40%) that are expressed during reprogramming are targets for NatA-mediated Nt acetylation. Of course, due to the limitation of the LC-MS itself, we expect some percentage of the proteome (with proteins expressed during reprogramming) was not successfully detected by this assay. But nevertheless, the observation of the enrichment of NatA substrates (detected) during reprogramming is consistent with the upregulated expression of both NAA10 and NAA15 (the catalytic and auxiliary components of NatA) as well as NAA50 and HYPK (the two proteins in the NatA complex) (Arnesen et al., 2006, 2010) seen from the same proteomics data shown in the above study (Hansson et al., 2012) (Fig 4.3 d). As a matter of fact, NatA is the only NATs that had an increased expression of all its components, suggesting a critical role that NatA might play during reprogramming of mouse embryonic fibroblasts (MEF) to iPSCs. However note that, the expression of NAA10 and NAA15 does not change stoichiometrically, suggesting a separate function of NAA10 (and NAA15) that might be independent of NatA activity (as discussed in chapter 3) during reprogramming.

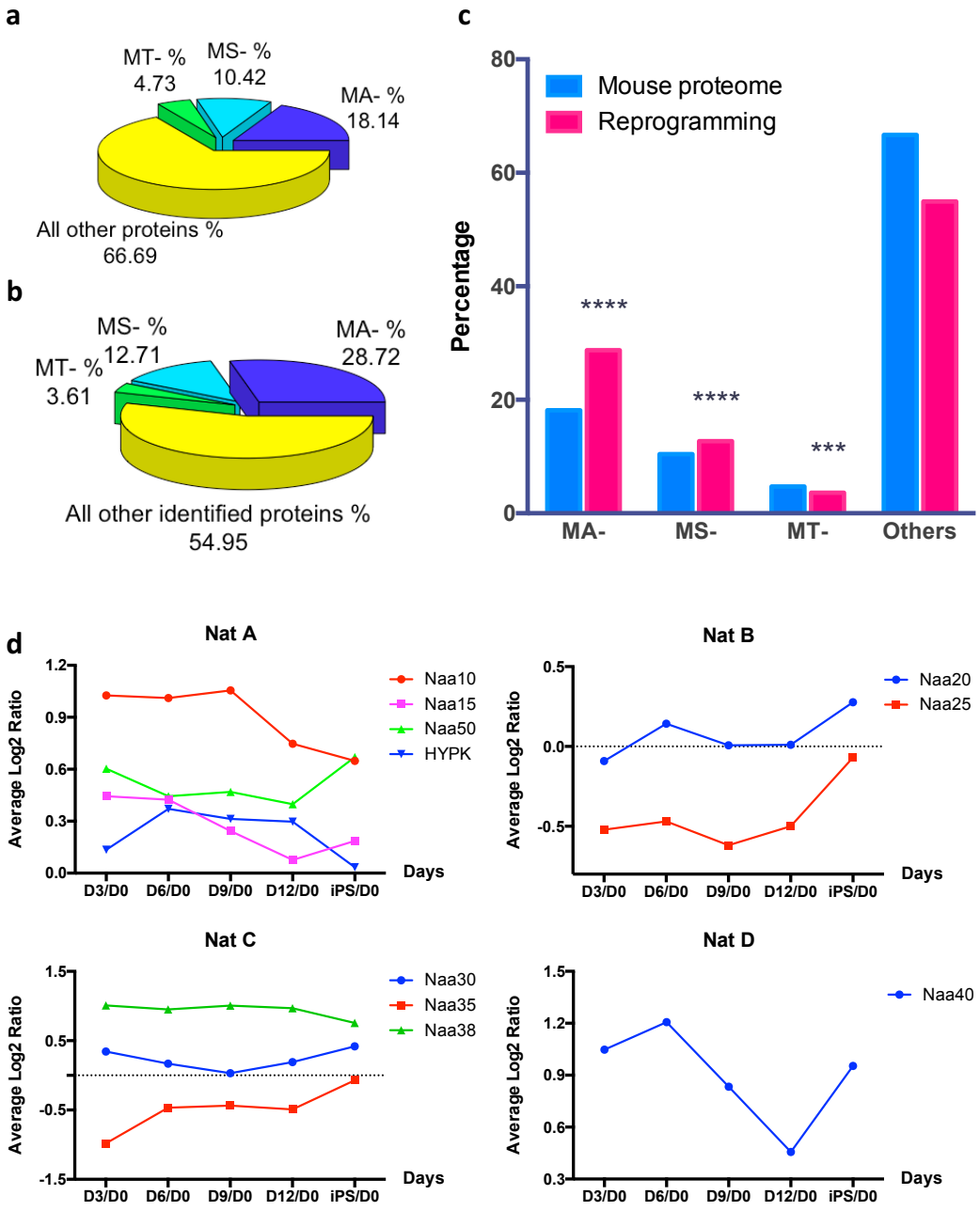


Figure 4.3 Profile of NatA substrates in mouse proteome during reprogramming. (a) An overall distribution of three major NatA substrate classes with indicated N-termini in the mouse proteome (UniProt release 2013_12). (b) Distribution of three major NatA substrate classes (MS-, MA-, MT-starting proteins) with indicated N-termini during reprogramming (data was from Hansson et al., 2012). (c) Bar plot comparison of a and b. ***$p < 0.001$; ****$p < 0.0001$. (d) The relative protein expression change of individual component of four NATs complexes (Nat A-D) during reprogramming of mouse embryonic fibroblasts (MEF) to iPSCs. Day 0, the beginning of the induction. Dn/D0, the ratio of protein expression on day N over day 0.

Generation of an iPSC model of Ogden syndrome

After five rounds of unsuccessful attempts to reprogram late passage number fibroblasts, we eventually obtained some early passage S37P (from one Ogden syndrome patient, Utah family III-6) and WT control fibroblasts (WT-1), and successfully derived iPSCs from these cells by introducing Sendai virus (SeV) based vectors (CytoTune-iPS 2.0 Sendai Reprogramming Kit, Thermo Fisher Scientific, #A16517) containing four reprogramming factors, OCT3/4, SOX2, C-MYC, and KLF4 (Takahashi, 2007). The Sendai virus-mediated reprogramming strategy is favorable in this study because of its unique merits of replicating exclusively in the cytoplasm of infected cells (H. O. Li et al., 2000) as well as being able to reprogram a broad range of somatic cells with a relatively high efficiency (Fusaki et al., 2009). For this particular reprogramming experiment, we did not see any obvious difference in reprogramming efficiencies between patient and control.

The workflow of the SeV-mediated reprogramming experiment is shown in (Fig 4.4 a). Both WT-1 and S37P iPSC colonies displayed a typical ESC-like colony morphology and they stained positive for hESCs surface antigenic marker TRA-1-60 and TRA-1-81 (Fig 4.4 b). We confirmed the clearance of the vectors and the exogenous reprogramming factor genes by qPCR after eight culture passages (data not shown). Upon the emergence of mature WT1- and S37P-iPSC colonies, three TRA1-60 positive stained iPSC colonies from each group (WT and S37P) were mechanically picked, transferred to feeder-free culture conditions, and propagated for line establishment.

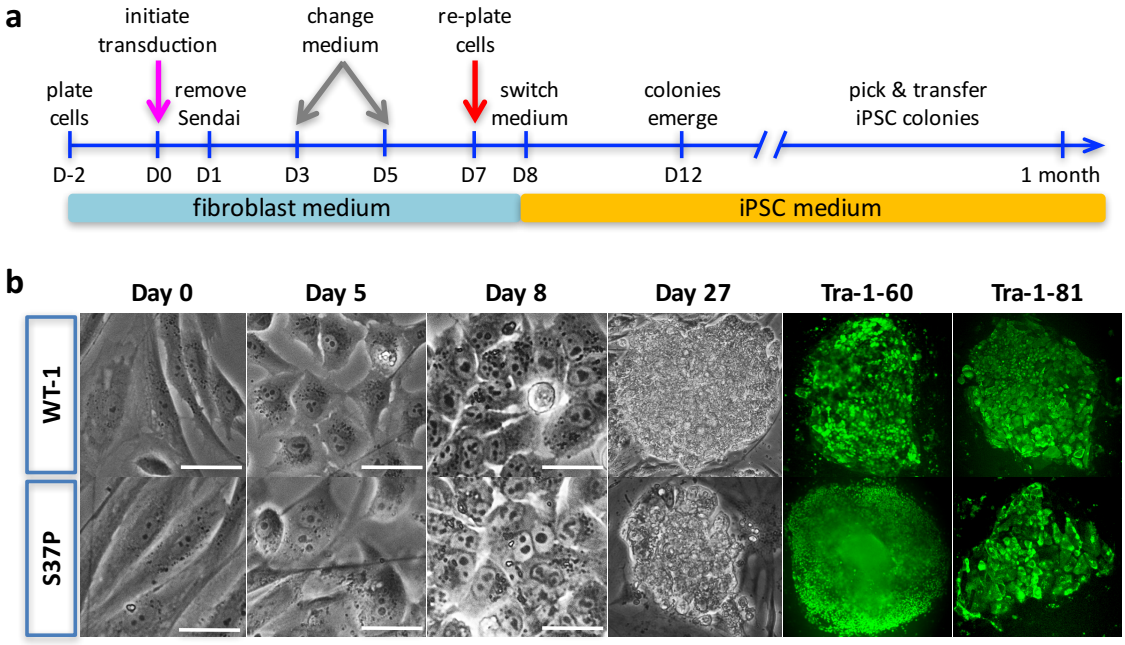


Figure 4.4 Generation of Sendai virus-induced pluripotent stem cells from primary fibroblasts. (a) Schematic depicting Sendai virus mediated reprogramming protocol and timeline of iPSC generation from skin fibroblasts. (b) Morphology changes of cells from mesenchyme to epithelial (MET) between day 0 to day 8 of reprogramming was monitored by bright field microscopy. Representative phase contrast images (20×) on Day 8 and Day 27 shown a high nuclear-to-cytoplasmic ratio and the appearance of hESC-like colonies respectively. Immunocytochemical analysis of pluripotency marker TRA-1-60 (images: 10×) on Day 26 and TRA-1-81 (images: 20×) on passage 5 iPSCs show positive staining patterns. Scale bar represents 50 µm. Cells pictures on day 0, 5, 8 in (b) were taken by Sunita D'Souza at Mt. Sinai.

Six iPSC lines were first genotyped to confirm the WT and Ogden syndrome alleles in *NAA10* (Fig 4.5 a). Then, the endogenous expression of the pluripotency associated transcription factors OCT4, NANOG, and SOX2 was evaluated by immunofluorescence analysis and revealed characteristics of pluripotent ES cells (Fig 4.5 b). Endogenous expression of *OCT4*, *NANOG*, *SOX2* and *KLF1* was also confirmed by real-time quantitative PCR (qPCR) (Fig 4.5 c) compared to two human embryonic stem cell (hESC) lines H1 (male derived) and H9 (female derived) (Thomson et al., 1998). Karyotyping was also performed and revealed chromosomal abnormalities in two of the WT iPSC lines (WT-1 #1 and #5) (Fig 4.5 d); hence, these two lines were eliminated from the downstream assays. In order to increase the statistical power of the experiment design, we purchased two more WT hiPSC lines (WT-4 & 6) which were both derived from BJ fibroblasts (hence age, race and gender matched) using Sendai virus mediated reprogramming method. These two hiPSC lines were already well characterized, hence they were not included in our assays above. During overall observations, S37P cells did not display any difference during reprogramming compared to the WT.

S37P-H embryoid bodies show defect in growth

To confirm three germ layer differentiation capacity of derived iPSCs, I performed a spontaneous embryoid body (EB) differentiation assay in a suspension culture. Both groups of EBs were formed without difficulty (Fig 4.5 e). However, very slow growth and a much higher cell death rate were seen from the patient's iPSCs-formed EBs (S37P-H iPSCs) during the time course of the experiment. By the end of a five-week culture, the size of S37P-H EBs were much smaller compared to the WT ones (Fig 4.5 e), which made it impossible for histology analysis. The histology study of the WT EBs revealed tissues specific to each of the three embryonic germ layers (Fig 4.5 f).

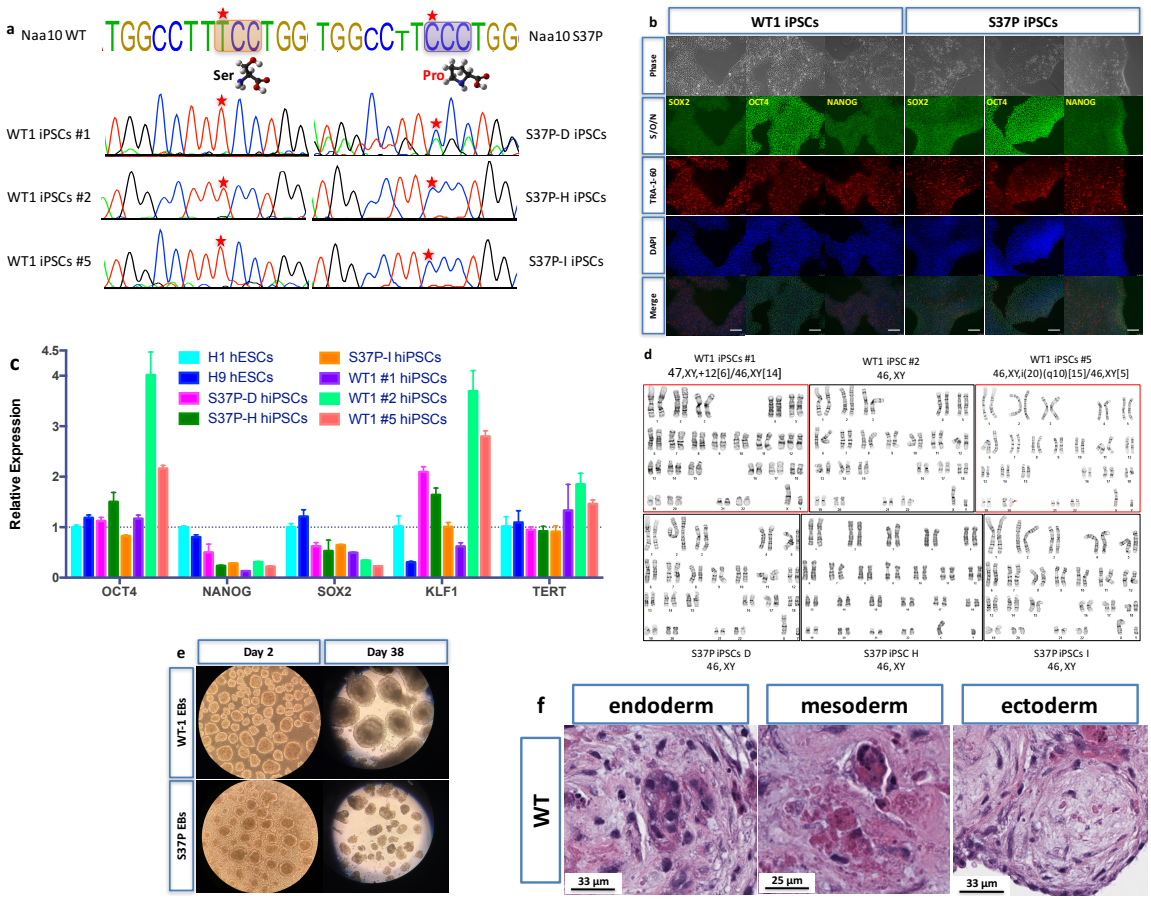


Figure 4.5 Characterization of human iPSC line WT-1 #1, #2, #5 and S37P-D, I, H. (a) DNA sequencing analysis of NAA10 confirmed the WT and c.109T>C alleles in WT1- and S37P-derived iPSCs respectively. The c.109 allele is indicated by the red asterisk, and the codon for p.S37 is highlighted. (b) Representative immunofluorescent staining for SOX2, OCT4, NANOG, and TRA-1-60 in WT1 and S37P derived iPSC colonies. Nuclei were stained with DAPI (blue). Bar indicates 150 µm. (c) Expression of *OCT4*, *NANOG*, *SOX2*, *KLF1*, and *TERT* were analyzed by quantitative real-time PCR in human ESC lines (H1, H9) and six derived iPSC lines. Samples were normalized against the internal control (beta-actin) and plotted relative to the expression level in the human ES cell line H1 (left-most column), which is arbitrarily set to 1, using the ΔΔCt method. Statistical analysis was not performed to compare the expression differences among lines. This experiment needs to be repeated prior to publication. (d) Karyograms of six iPSC lines studied. Abnormal karyotypes are squared in red. (e) Suspension culture of embryoid bodies (EBs) formed from WT-1 and S37P-H iPSCs at day 2 and day 38. (f) Representative images of *in vitro* differentiation of WT-1 iPSCs showing tissues of three germ layers; neuroglial cells sitting in a fibrillary schwannian stroma (ectoderm), hematopoietic lineage cells (mesoderm) and a glandular structure (endoderm). EB assay was repeated three times with EBs formed from iPSCs of three different passage numbers. Karyotyping experiments were carried out by commercial services. Sunita D'Souza and Vera Alexeeva at Mt. Sinai iPSC CORE assisted with experiments shown in b and c, respectively.

The reasons for defective growth of EBs formed from S37P-H iPSCs could be many, but in order to have a better picture and draw a definitive conclusion, repeats of EB formation assay using other S37P iPSC lines (from the same patient and from other patients if ignoring the accessibility issue) needs to be carried out first. Meanwhile, it would be also helpful and informative to use different protocols to evaluate the trilineage differentiation potential of the patient-derived iPSCs, such as promoting the differentiation in an adherent culture following the immunocytochemistry analysis or examining the gene expression of the three germ layer markers in two-week old EBs. Furthermore, one could also carry out the teratoma formation assay and/or tetraploid complementation assay to test the pluripotency of iPSCs *in vivo*. If the EBs formed from other S37P iPSC lines can grow and differentiate without issues, it might just indicate a low quality of this S37P-H iPSC line. If all S37P iPSC lines replicate the same defect in growth and differentiation, there is a higher chance that this phenotype is disease-relevant.

One possible explanation of the failed S37P EB growth in suspension culture (besides the quality concern) is the slow growth of S37P iPSCs within a 3D structure versus the rapid cell death under an extended *in vitro* culture. On the one hand, as seen in the cell density experiment in Chapter 3, S37P cells can adjust their proliferation and viability to respond to different culture densities. Hence, it is quite possible that they may acquire different dynamics of growth in a 3D structure. On the other hand, it has been a consistent observation that S37P cells have a vulnerability to stress-induced cell death under unfavorable conditions such as an extended culture. As mentioned in the discussion section of Chapter 3, NAA10/NatA regulates cell proliferation and survival through directly interacting with proteins that have influences on autophagy and apoptosis pathways.

Directed differentiation of hiPSCs into cardiomyocytes

We differentiated WT-1 and S37P iPSCs into cardiomyocytes (iPS-CMs) by temporally modulating WNT/β-catenin signaling to investigate the cardiac phenotype of Ogden syndrome. Studies have shown that Wnt signaling has a biphasic effect on heart development with early enhancing and later inhibiting cardiomyogenesis (Naito et al., 2006; Ueno et al., 2007). Our differentiation protocol was adapted from the "GiWi" protocol (Lian, Hsiao, et al., 2012; Lian, Zhang, et al., 2012), which applies two small-molecule regulators of canonical Wnt signaling, Gsk3 inhibitor (CHIR) and Wnt inhibitor (IWP2), at precise developmental stages of human pluripotent stem cells (hPSCs) to sequentially promote mesoderm formation and cardiomyocyte specification (Fig 4.6 a).

Using this protocol, I efficiently generated functional cardiomyocytes from WT-1 and S37P iPSCs (P20~P25) on both plastic and glass bottom plates (coated with Matrigel) in a completely defined, growth factor- and serum-free system (Fig 4.6 b). We did not perform a full evaluation of the cardiomyocytes yield from this protocol; however, an initial flow cytometry analysis of the cardiac troponin T (cTnT) expression from one month old WT-1 iPS-CMs culture revealed an overall differentiation efficiency above 90% (date not included due to this being a single experiment). But a comparable differentiation efficiency from S37P iPS-CMs can be estimated based on the percentage of the spontaneous beating cells in the culture. Spontaneous and synchronous contractions from the cell population on 12-well plates were typically noticed around 12 days after the initiation of differentiation, and can be maintained in culture for more than six months. Notably, with the extended time of in culture, iPS-CMs started to form 3D functional syncytia, which is consistent with previous published findings and is believed to give them a more mature phenotype (Hirt et al., 2014). Also, when iPSCs were induced on 96-well

plates (glass bottom), they tended to not only have a higher differentiation efficiency, but also had spontaneous beating activity as early as day 7, which could be possibly due to a number of reasons including a steady concentration (less evaporation) and a more uniform distribution of the small molecules used for the induction. Immunofluorescence microscopy analysis showed that iPS-CMs at five-month old had protein-expression profiles consistent with a cardiac lineage (Fig 4.7). We did not perform any quantification of gene or protein expression for these cardiac specific markers.

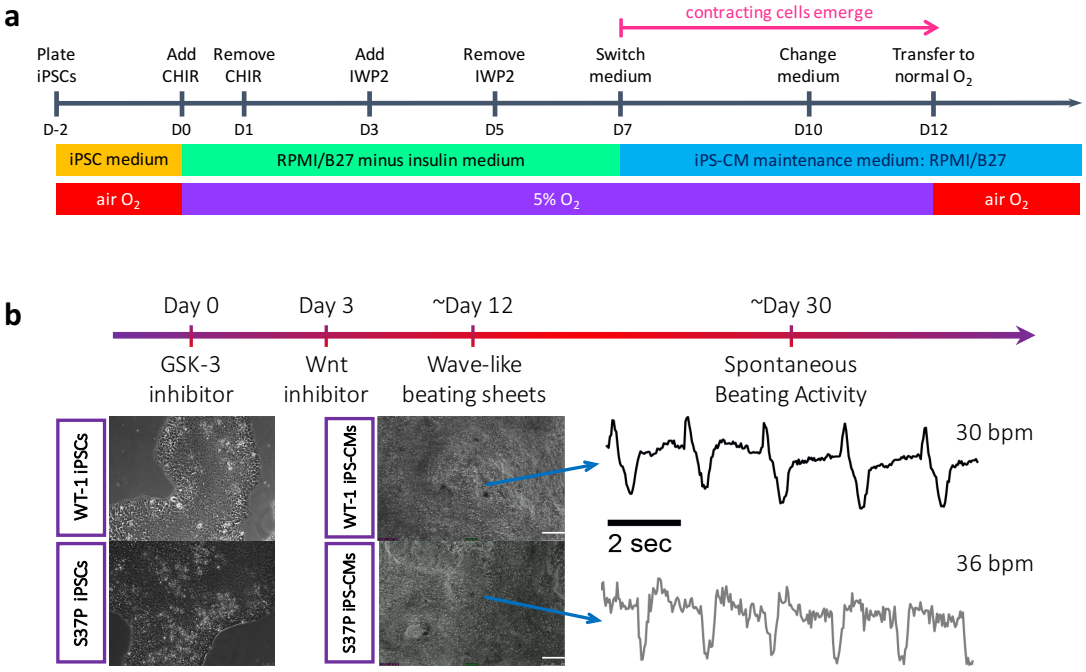


Figure 4.6 Generation of iPSC-derived cardiomyocytes using "GiWi" protocol. (a) Schematic depicting cardiomyocytes differentiation protocol by modulating WNT/β-catenin signaling from iPSCs. (b) Morphology change of cells during differentiation was monitored by bright field microscopy. Representative phase contrast images (10×) shown a typical human ESC-like iPSC colony under feeder-free culture condition on Day 0 of induction and the appearance of a thicker, wave-like beating sheet on ~day 40 of induction. Spontaneous beating activities of iPS-CM were recorded and traces were extracted at a month. Statistical testing of the difference in baseline beating rates between groups was not performed. Scale bar, 250 μm. Spontaneous beating activity traces in (b) were extracted and plotted by Aleks Klimas from the videos that I recorded.

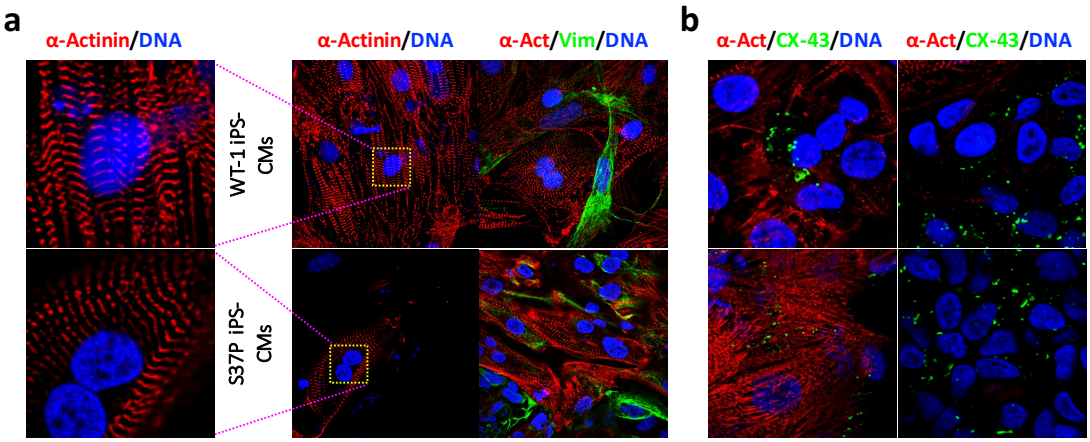


Figure 4.7 Immunostaining analysis of iPS-CMs at 5 months old. Images show iPS-CMs stained with (a) sarcomeric α-actinin (red), fibroblasts marker vimentin (green), and nuclei stain DAPI (blue), images: 100×. (b) α-actinin (red), the major gap-junction protein connexin 43 (CX-43, green), and DAPI (blue), images: 200×.

The differentiation capacity of S37P-iPSCs into cardiomyocytes did not appear to be affected by the mutation compared to the WT controls; however, with extended time in culture, we saw relatively more cell death from the patient iPS-CMs compared to the WT controls. The vulnerability to stress of S37P cells seems to be a consistent finding among all our experiments. We characterized our iPS-CMs at both early-stage (1 month) and late-stage (3-6 months), since many studies have shown that a prolonged time in culture can help enhance the maturation of iPS derived cardiomyocytes *in vitro* toward adult heart (Lundy et al., 2013; Veerman et al., 2015; Yang et al., 2014). Hence, studying late-stage iPS-CMs can help us understand the progression of the phenotype along time. But at the same time, our Ogden syndrome patients were infants before passing away, thus it is equally important to collect data at early-stage for a comparison purpose.

S37P iPS-CMs show frequent beating irregularity

The spontaneous baseline rhythms from both the patient and WT iPS-CMs recorded at a month of induction were between 30 to 40 bpm (Fig 4.6 b), which is typically seen in iPS-CMs (Reppel et al., 2004; Xu et al., 2002; Zhang et al., 2009), although note that we did not record enough baseline activities for statistical testing. During routine cultures and electrophysiology recording experiments, we noticed S37P iPS-CMs often presented spontaneous arrhythmic activities, such as bigeminy-like/missed beats (a), complex irregularity (b), and extra beats (c) (Fig 4.8). These rhythms have not been seen from the WT iPS-CMs either generated by us or the commercial ones we cultured in the lab, but they reflected what was documented in Ogden syndrome patients, who frequently presented all types of cardiac dysrhythmias.

A clear striated pattern of sarcomeric organization in S37P iPS-CMs revealed by a Z-line protein alpha-actinin staining did not reveal any obvious differences compared to controls. The distribution of cardiac fibroblasts which stained positive for vimentin also seemed to be comparable between the patient and WT iPS-CMs (Fig 4.7 a). Furthermore, the key gap junction protein expressed between cells in cardiac clusters Connexin 43 (Cx-43) was also detected in iPS derived cardiomyocyte culture, indicating the presence of functional electromechanical coupling (Fig 4.7 b). Notably, this cell-cell junction molecule not only was identified between CM-CM pairs, but also between fibroblast-fibroblast (Fb-Fb) pairs as well as between Fb-CM pairs. This observation has been also documented by other studies, indicating the important role of cardiac fibroblasts in modulating cardiac myocyte electrophysiology (Gaudesius et al., 2003; Goldsmith et al., 2004; Zhang et al., 2012). In fact, we observed more Cx-43 expression localizing to the perinuclear region of non-myocytes in our iPS-CMs culture (for both WT and S37P) than a circumferential distribution between CM-CM pairs, which was confirmed in commercial hiPSC-derived cardiomyocyte culture too (Axiogenesis, data not shown). The reason is unclear to us, and could be due to some remodeling process during such a prolonged in culture (about 6 months), but relevant data of the characterization of long-term cultured iPS-CMs is not available for comparison.

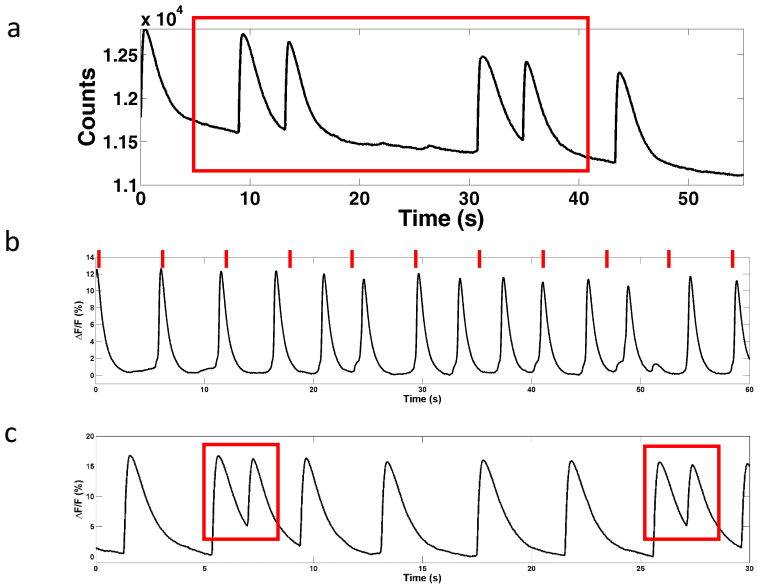


Figure 4.8 Spontaneous arrhythmic activities of 3 months old S37P iPS-CMs. Traces extracted from intracellular calcium transients recording under electrical stimulation show (a) bigeminy-like/missed beats, (b) complex irregularity, and (c) extra beats. I pre-treated the iPS-CMs (derived by me) with the calcium indicator dye Rhod-4, and Aleks Klimas performed the following electrophysiology studies, and analyzed and plotted the data shown.

Functional characterization of WT and S37P iPS-CMs

All electrophysiology (EP) studies of iPS-CMs were carried out using OptoDyCE system, a high-throughput all-optical cardiac electrophysiology platform developed at Dr. Entcheva lab (Burton et al., 2015; Klimas et al., 2016). iPSC-CMs subjected to all-optical phenotyping were first optogenetically transformed to expressed channelrhodopsin-2 (ChR2(H134R)) (Fig 4.9). Our project is the first application of OptoDyCE in studying patient derived iPS-CMs. The unique ability to perform optical pacing at multiple frequencies using this all-optical platform provides us an efficient tool to pinpoint the sources of cardiac arrhythmias in Ogden syndrome patients.

All EP results discussed below are based on evaluations of iPS-CMs derived on plastic bottom 12-well plates later being re-plated on glass bottom 14 mm dish for characterization. Aleks also helped perform the EP studies of iPS-CMs derived directly on glass bottom 96-well plates (made by myself) together with mexiletine (Na^+ channel blocker) treatment; however unfortunately, that data has yet to be analyzed to be discussed in this dissertation.

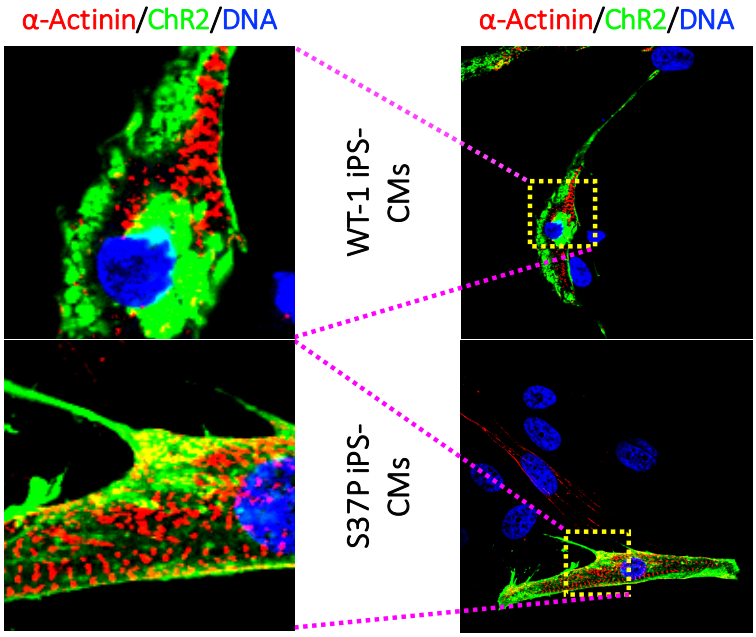
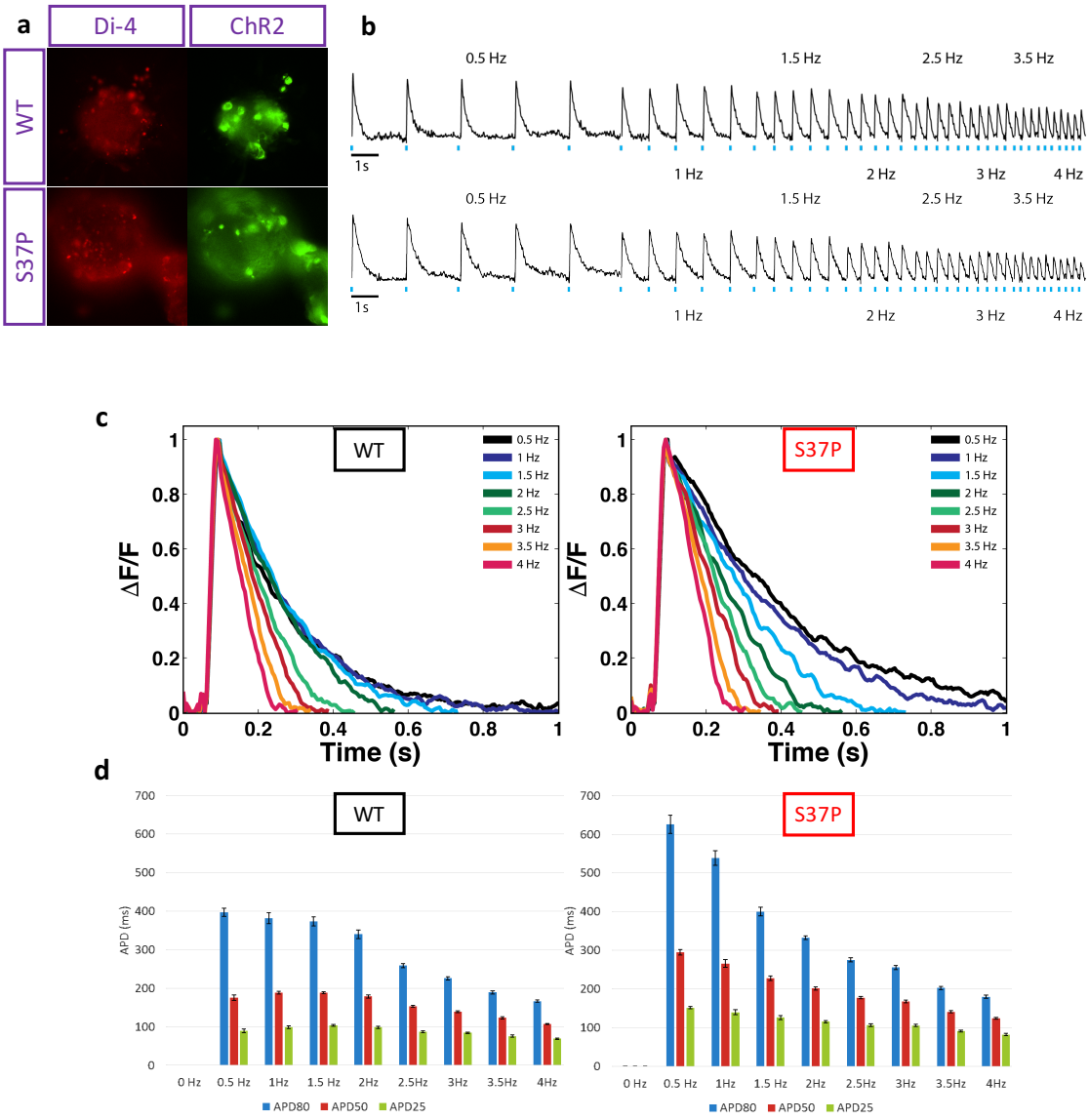


Figure 4.9 Optogenetically modified iPSC-CMs express ChR2. 6 months old iPS-CMs, immunostained for sarcomeric α-actinin (red) and nuclei stain DAPI (blue), were genetically modified via an adenoviral delivery system to express channelrhodopsin-2 (ChR2(H134R)) (eYFP) to enable optogenetic actuation by OptoDyCE system. Images (right column): 100×. Aleks Klimas helped with the addition of the virus to iPS-CMs.

First, we evaluated the action potentials (APs) of both patient and WT iPS-CMs. The optogenetically transformed iPSC-CMs were labeled with synthetic voltage-sensitive dye Di-4 and expressed ChR2 (Fig 4.10 a). APs were optically recorded at pacing frequencies from 0.5 Hz to 4 Hz to evaluate the restitution properties (Fig 4.10 b). APs recorded from both WT and patient samples shared a very similar morphology (Fig 4.10 c, e), and this morphology was dominant for both groups. Prolonged averaged APs recorded from the patient's iPS-CMs were noticed compared to the control samples but only at lower pacing frequencies (Fig 4.10 c, e), which is consistent with imbalance in depolarization and hyperpolarizing currents seen in prolonged QT intervals (Abrams & Macrae, 2014). We then quantified and compared action potential durations (APDs) measured at 25, 50 and 80% repolarization (APD25, APD50 and APD80, respectively), and a significant difference was seen between the patient and control samples at 0.5 Hz and 1 Hz pacing frequencies (Fig 4.10 d).

The restitution curves using APD25, APD50 and APD80 data from the patient and control iPS-CMs were then constructed. First introduced in 1968 and defined in 1975, electrical restitution curve (ERC) of action potential, which can be given by plotting the action potential duration (APD) as a function of preceding diastolic interval (DI) or cycle length (CL) between a steady-state response and an extrastimulus response respectively, provides a first approximation of the system's dynamics (Bass, 1975; Elharrar & Surawicz, 1983; Nolasco & Dahlen, 1968). A flattening ERC is protective, whereas a steep ERC indicates the electrical instability, i.e. the vulnerability to arrhythmia (FRANZ, 2003; Gelzer et al., 2008; Sabir et al., 2008). As seen in Fig 4.10 f, the patient sample presented a much steeper ERC of APD compared to the control, and a statistically significant prolongation of APD was observed at 0.5 Hz between patient and control iPS-CMs. All differences seen between S37P and WT samples were consistent no matter how

the data set was pooled and compared (i.e. different dates, different FOVs, different clones). Together, these findings from AP recording further proved that heart cells of Ogden syndrome patients have a poor cardiac electrical stability and are arrhythmogenic especially at a low heart rate.


a
Di-4
ChR2
WT
S37P
b
0.5 Hz
1.5 Hz
2.5 Hz
3.5 Hz
1s
1 Hz
2 Hz
3 Hz
4 Hz
c
WT
S37P
ΔF/F
Time (s)
0.5 Hz
1 Hz
1.5 Hz
2 Hz
2.5 Hz
3 Hz
3.5 Hz
4 Hz
d
APD (ms)
0 Hz
0.5 Hz
1Hz
1.5 Hz
2Hz
2.5Hz
3Hz
3.5Hz
4Hz
APD80
APD50
APD25

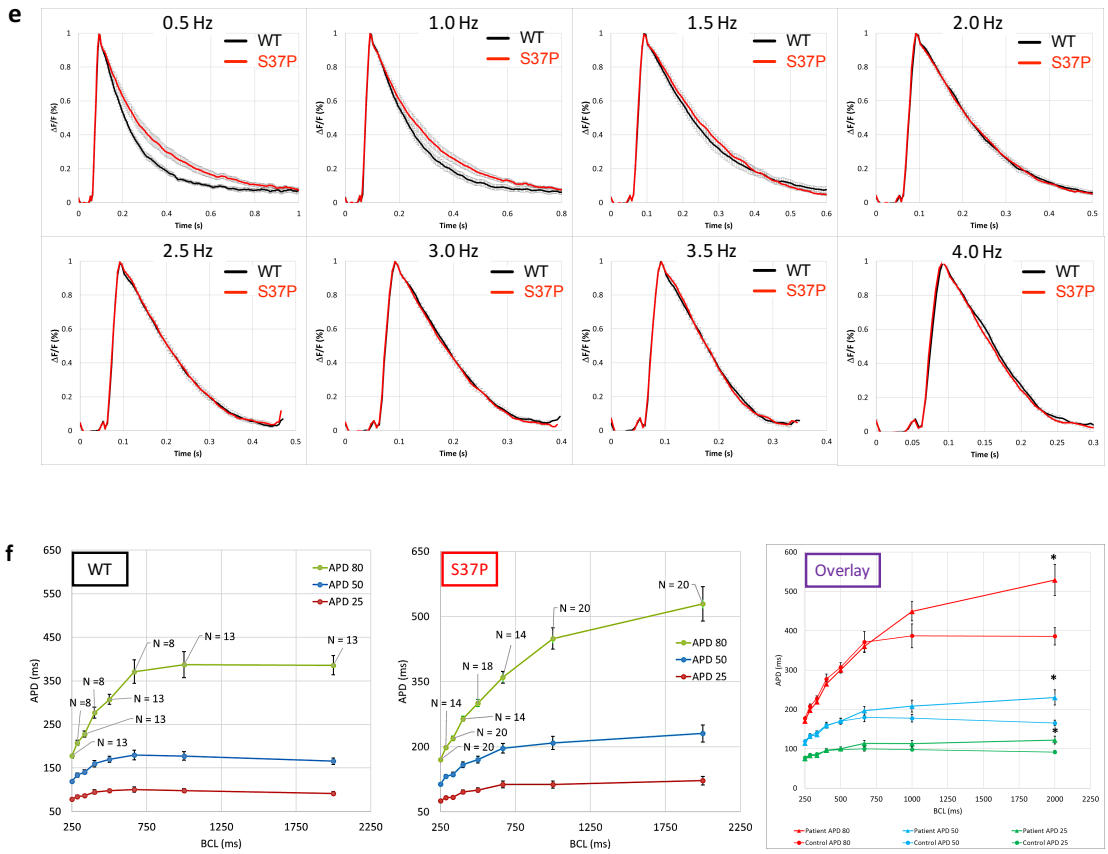


Figure 4.10 Membrane potential recording of 5- and 6-month old iPS-CMs by optical stimulation. (a) Optogenetically transformed iPSC-CMs was labled with synthetic voltage-sensitive dye Di-4 (red) and express ChR2 (eYFP, green). (b) Example of membrane potentials from WT and patient iPS-CMs responding to 0.5 Hz and 4.0 Hz optical pacing frequencies. (c) Averaged membrane potential traces under all pacing frequencies overlaid within each group. (d) The averaged APs were quantified by 3 parameters; AP duration (APD) at 25, 50 and 80% repolarization (APD25, APD50 and APD80, respectively). (e) Averaged membrane potential traces, given terms of ΔF/F (%), for patient (red) and WT (black) at multiple pacing frequencies. (f) Individual and overlaid APD restitution curves using all data combined. Sample size (field of view) for each measurement is indicated with the N number. All quantitative results are represented as mean ± standard error of the mean. *$p < 0.05$ using 2-way ANOVA with Tukey-Kramer post-hoc test for pacing at 0.5Hz for APD25/50/80 between patient and control groups. Replicates and repeats for experiments in d-f under different pacing frequencies were stated in figure (f). WT data is from WT-1 #2 iPS-CMs, and S37P data is from the combination of S37P-D and H iPS-CMs. All experiments were repeated three times. I pre-treated the iPS-CMs (derived by me) with the synthetic voltage-sensitive dye Di-4-ANBDQBS, and Aleks Klimas performed the following electrophysiology studies, and analyzed and plotted the data shown. Aleks Klimas also helped with the addition of the virus to iPS-CMs.

For the intracellular calcium ($[Ca^{2+}]_i$) transient evaluation, we first performed recordings using 3 months old iPS-CMs by electrical stimulation. As mentioned above, spontaneous arrhythmic acitivities were commonly seen in S37P iPS-CMs. In order to avoid the confounding interference by the intrinsic arrhythmic activities from the patient iPS-CMs, we usually preferred to select areas that had no spontaneous beating activities for recording. But when we did encounter such spontaneous rhythm in the patient's sample during recording, we usually found it subsided when external pacing was delivered to the cells, especially when the pacing frequency was higher than the intrinsic one but could still be followed by the cells (Fig 4.11). This probably can be mainly, if not entirely, explained by the mechanism called overdrive suppression (Vassalle, 1977).

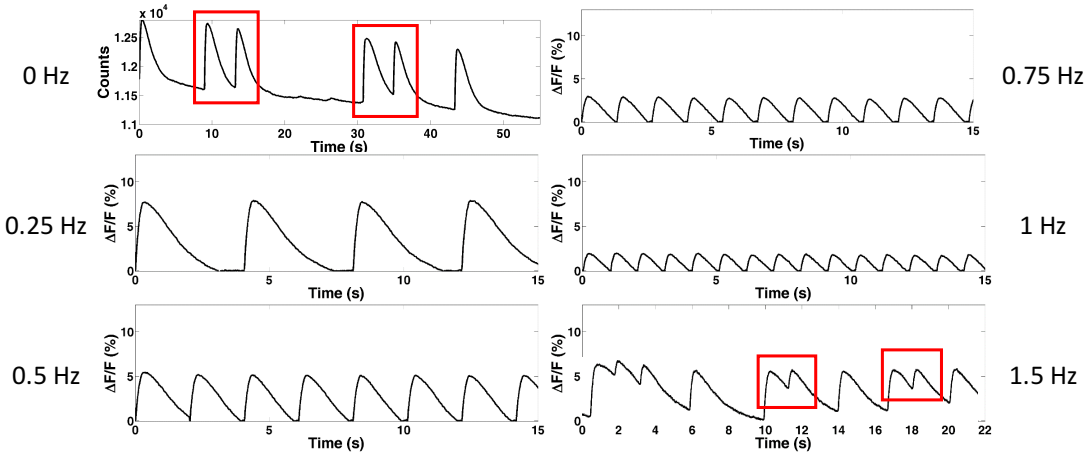


Figure 4.11 Spontaneous arrhythmic activities of 3-month old S37P-H iPS-CMs during calcium transient recording. Traces extracted from calcium transients recording of 3M S37P iPS-CMs show extra beats under electrical stimulation at 0 Hz and 1.5 Hz frequencies. I pre-treated the iPS-CMs (derived by me) with the calcium indicator dye Rhod-4, and Aleks Klimas performed the following electrophysiology studies, and analyzed and plotted the data shown.

For multiple field of views studied during electrical pacing, both WT and S37P iPS-CMs can follow up to 1 Hz of external pacing (Fig 4.12 a, c). When averaged calcium transients were compared, we observed a reduced upstroke velocity (representing the release of Ca^{2+} from the sarcoplasmic reticulum through ryanodine receptor channels (Bers, 2002)) in patient cells compared to the WT control (Fig 4.12 b), and it was further proved by quantifying the time to reach peak height (time to peak) at each pacing frequency (Fig 4.12 d). Such differences, though not significant upon statistical testing (most likely due to lack of power), were consistent across all pacing frequencies and especially evident at a higher pacing frequencies. Lastly, we analyzed the restitution properties of calcium transient durations, which did not reveal obvious difference between patient and control (Fig 4.12 e).

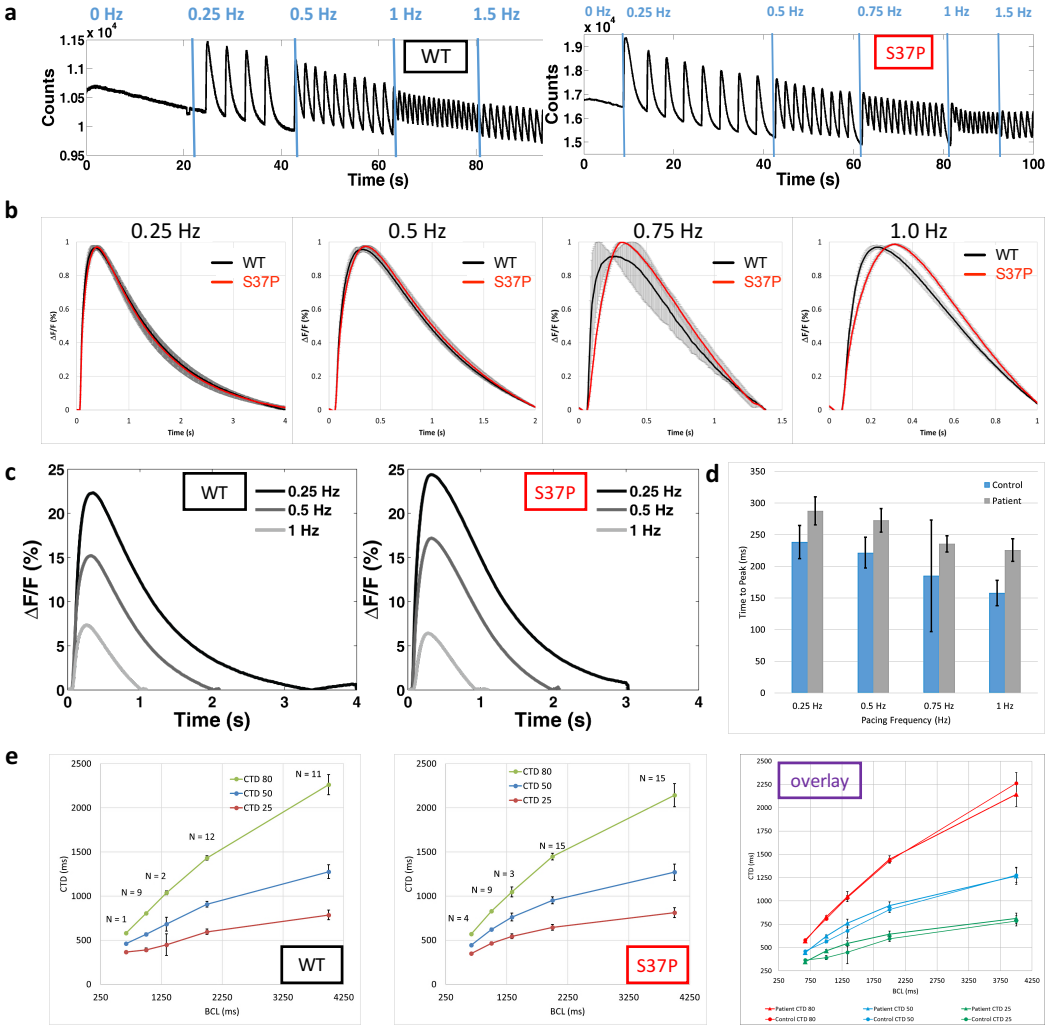


Figure 4.12 Intracellular calcium transient evaluation of 3-month old iPS-CMs by electrical stimulation. (a) Example of intracellular calcium transients of iPS-CMs responding to 0.25 Hz to 1.5 Hz pacing frequencies. (b) Averaged calcium transients from WT and patient iPS-CMs at multiple pacing frequencies. (c) Example of calcium transients under all pacing frequencies overlaid within each group. (d) Comparison of time to peak between WT and patient. (e) Individual and overlaid calcium transient restitution curves using all data combined. The averaged CTs were quantified by 3 parameters; CT duration (CTD) at 25, 50 and 80% return (CTD25, CTD50 and CTD80, respectively). Sample size (field of view) for each measurement is indicated with the N number. All quantitative results are represented as mean ± standard error of the mean. Replicates and repeats for experiments in (b, d, e) under different pacing frequencies were stated in figure (e). WT data is from WT-1 #2 iPS-CMs, and S37P data is from the combination of S37P-D, H, I iPS-CMs. All experiments were repeated three times. I pre-treated the iPS-CMs (derived by me) with the calcium indicator dye Rhod-4, and Aleks Klimas performed the following electrophysiology studies, and analyzed and plotted the data shown.

We later repeated the calcium transients recording in six months old iPS-CMs using optical pacing, and the expression of channelrhodopsin-2 (ChR2) together with a positive staining for the Rhod4 dye were confirmed in the recorded cells (Fig 4.13 a). However, most of the calcium transient (CT) signals recorded unfortunately had a high noise level (Fig 4.13 b). Together with the fact that we only successfully obtained CT data from one clone of each group, I do not think the result is reliable enough to be included in this dissertation.

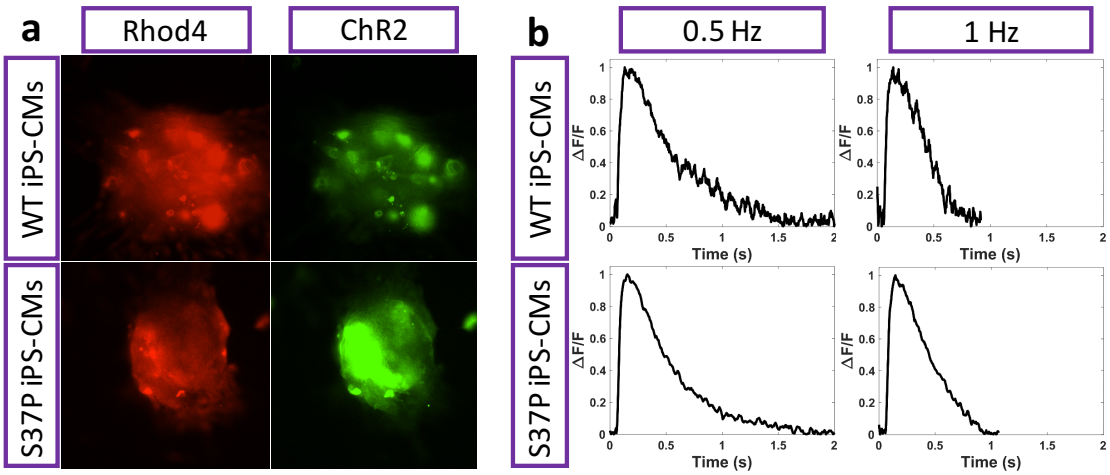


Figure 4.13 Intracellular calcium transient evaluation of 6-month old iPS-CMs by optical stimulation. (a) Optogenetically transformed hiPSC-CMs was labled with calcium indicator dye Rhod-4 (red) and express ChR2 (eYFP, green). (b) Example of averaged intracellular calcium transient signals from WT and patient iPS-CMs responding to 0.5 Hz and 1.0 Hz optical pacing frequencies. Motion artifacts were seen in all figures, especially in WT traces. WT data is from WT-1 #2 iPS-CMs, and S37P data is from S37P-H iPS-CMs. This experiment was carried out on a single day with 3 FOVs recorded from the WT-1 iPS-CMs and 9 FOVs recorded from S37P-H iPS-CMs. I pre-treated the iPS-CMs (derived by me) with the calcium indicator dye Rhod-4, and Aleks Klimas performed the following electrophysiology studies, and analyzed and plotted the data shown. Aleks Klimas also helped with the addition of the virus to iPS-CMs.

Lastly, as waiting for the new batch of EP study data from recording 1M iPS-CMs derived directly on glass bottom 96-well plates to be analyzed by Aleks, our hope is to not only validate our observations above, but also to shed some new light on mechanisms underlying these observations, i.e. understand more of Na^+ channel property in S37P cells based on the mexiletine treatment experiment. Also, being cultured for a much shorter time *in vitro*, these cells had a high monolayer morphology before recording, which facilitated the measurement further. However, in the meantime, since CMs studied in this batch of experiments are very young compared to our previously analyzed samples, differences will be also expected due to different degrees of maturity of the cells.

4.4 Discussion

Protein Nt-acetylation in heart disease pathogenesis

The action potential prolongation only revealed at longer pacing cycle lengths in S37P iPS-CMs is very similar to what was found in type 3 Long QT syndrome (LQT3) patients, who are more likely to develop spontaneous arrhythmic events (i.e. torsades de pointes) at slow heart rates (i.e. rest, sleep) rather than as a consequence of sympathetic activation (rapid heart rates, i.e. exercise) (Bohnen et al., 2017; Schwartz et al., 1995). LQT3 is caused by gain-of-function mutations in SCN5A (Abriel & Zaklyazminskaya, 2013), encoding the dominant voltage-gated sodium channel in the heart, $Na_v1.5$. In LQT3, a fraction of defective Na^+ channels fail to inactivate (burst mode), which leads to sustained macroscopic inward Na^+ channel current that can delay repolarization and prolong the QT interval at slow heart rates (Clancy & Rudy, 1999; Clancy et al., 2002). $Na_v1.5$ plays a critical role in rapidly initiating and efficiently propagating the action potentials in the cardiac tissue (Kléber & Rudy, 2004). Malfunctions of $Na_v1.5$

function, either caused by mutations in SCN5A or secondary to cardiac disorders, can give rise to many types of arrhythmias that lead to sudden cardiac death (John et al., 2012). Studies have shown that $Na_V1.5$ is subject to a wide range of post-translational modifications (PTMs) under both physiological conditions and diseases, including N-glycosylation, arginine methylation, redox modification, ubiquitination, S-nitrosylation, and phosphorylation by various kinases such as protein kinases A and C, Ca^{2+}/Calmodulin-dependent protein kinase II, phosphatidylinositol 3-kinase (Marionneau & Abriel, 2015).

Among all the PTMs of $Na_v1.5$, the most recent discovery of the role of N-terminal acetylation of $Na_v1.5$ in end stage heart failure (HF) is of particular interest (Beltran-Alvarez et al., 2014), since this is the first and the only evidence to date for NTA in regulating ion channel proteins and heart disease pathogenesis in general. In this report, researchers discovered that the Nt-acetylated $Na_V1.5$ is the most abundant (if not only) $Na_V1.5$ form in the heart tissues from HF patients, and NTA helps target the protein for ubiquitin-dependent degradation. Also, the authors speculated that NTA of $Na_V1.5$ might be increased in end-stage HF, due to the fact that a reduction of $Na_V1.5$ expression is typically seen in HF patients' cells and N terminus of $Na_V1.5$ in healthy mice's heart tissues is intact (including the initiation methionine). On the other hand, other recent research on end-stage failing human heart demonstrated an extensive mitochondrial protein Nε-lysine hyperacetylation (Horton et al., 2016), among which many are substrates of NatA. It would be interesting to perform a Nt-acetylproteome analysis using the same samples and determine whether there is any relationship between these two protein modifications to hint of some crosstalk.

In theory, $Na_V1.5$ is a classic substrate of NatA since the amino acid sequence of its N terminus is "$MA_{NFLLPRGT}$". The observation stated in above paper provides a piece of evidence

for an earlier hypothesis of mine that there might be a temporal and spatial regulation of Nt-acetylation on targeted proteins during development and disease pathogenesis. In fact, this idea has been recently supported by a few other scholars (Silva & Martinho, 2015). Though the biochemical reaction of NTA itself is considered to be irreversible and there is no evidence found to date supporting the existence of N-terminal deacetylases, other regulations of NTA *in vivo* can still be possible such as through regulation of NATs or adjusting acetyl-CoA levels. Also, if more evidence can support the role of protein NTA in disease pathogenesis, efforts can be put to develop NATs inhibitors *in vitro* as novel therapeutics to control or even reverse disease progress, just as in the field of histone acetylation where more than 20 of histone deacetylase inhibitors have been developed and used in treating various diseases including cancer (Qiu et al., 2017).

Insights of iPS-CMs disease modeling

The creation of an iPS-CM model for studying Ogden syndrome has brought us many interesting and valuable insights about this unique tool and process. Among them, some are challenges that have been well recognized in the field, and others are novel observations that need to be explored more to advance the field.

Besides recording APs from both 5-6 months old iPS-CMs using optical stimulation, we also performed the AP recording in 3-month old iPS-CMs by electrical pacing. But unfortunately, due to signal-to-noise ratio (SNR) problems and issues with motion artifacts from recording big 3D cell clusters (due to long time culture and lack of efficient cell dissociation protocol then), we obtained too few good APs to compare (only one from the patient group). However interestingly, we noticed that unlike the uniform morphology seen from 5-6 months old iPS-CMs, APs

recorded from cells at this age had a variety of morphology forms (Fig 4.14) with a wide range of APDs. In fact, AP morphology variability is not new to hiPS-CMs and even hESC-CMs researchers (Blazeski et al., 2012; Zhu et al., 2016), it used to be considered due to a mixture of subpopulations of CMs in the culture including nodal-, atrial- or ventricular-like CMs, which were usually differentiated by APD length (Peng et al., 2010; Zhang et al., 2009). But several newer studies proved that might not be the entire case. For example, a study suggested that iPSC-CMs (iCell Cardiomyocytes from Cellular Dynamics International) possess a unique AP morphology, and seeding densities can affect it greatly with no evidence for specific subpopulations (Du et al., 2015). Other research comparing APs from cells of 7-95 days old demonstrated that time in culture does affect the morphology, but still a variety of APs are seen within a specific developmental stage of iPS-CMs based on (Ben-Ari et al., 2016). But many studies including the above two tend to all believe that different functional maturation of CMs is the key to give rise to a variety of AP morphologies seen in culture. And an overall shift from nodal phenotype dominance in young cultures to ventricular phenotype dominance in older cultures was observed (Ben-Ari et al., 2016). That probably explains what we observed in our experiments too, with multiple AP morphologies seen in 3-month old CMs compared to an almost uniform AP morphology from 5-6 months old iPS-CMs, which represents a high purity of matured cells.

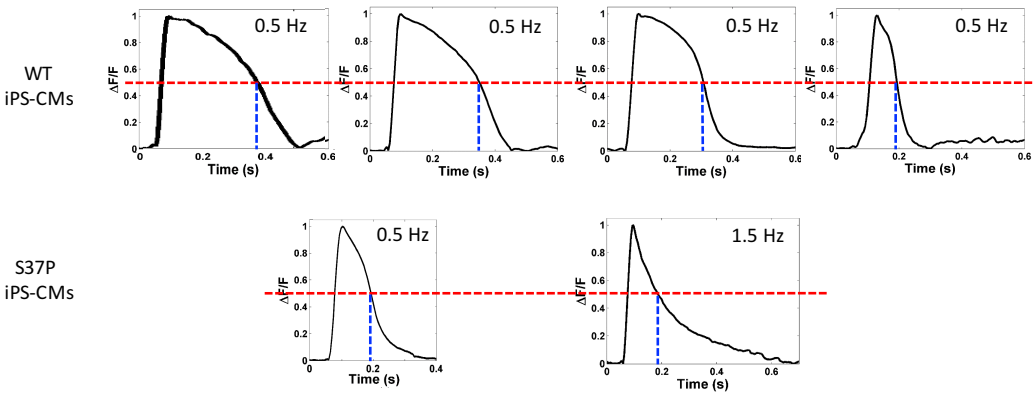


Figure 4.14 Varying membrane potential durations of 3-month old iPS-CMs by electrical stimulation. Averaged membrane potentials of 3M WT (4 FOVs under 0.5 Hz pacing) and S37P iPS-CMs (1 FOV, 2 frequencies) shown multiple AP morphologies. I pre-treated the iPS-CMs (derived by me) with the synthetic voltage-sensitive dye Di-4-ANBDQBS, and Aleks Klimas performed the following electrophysiology studies, and analyzed and plotted the data shown.

Admittedly, it is not practical to ask all researchers to wait such an extended period of time for the maturation of iPS-CMs to be maximal before any characterization. In fact, the most common practice in the field seems to perform the functional characterization of iPS-CMs at a month or two of induction, which can be potentially problematic if the data is generated based on a heterogeneous culture. Hence, there are increasing calls for establishing highly efficient differentiation protocols to yield a large number of relatively purified populations of CMs, or to expedite the *in vitro* maturation process such as by recreating the cardiac niche (Atmanli & Domian, 2016), or for developing reliable CMs subpopulation markers to enable an enrichment process. Moreover, as far as using iPS-CMs for disease modeling, more and more evidence needs to be collected to illustrate the correlation between the iPS-CMs age in culture and their corresponding age *in vivo*. The knowledge gained in this regard will help researchers choose the specific endpoints to investigate the phenotype to be around the disease onset age, especially when the disease of study is progressive and the phenotype does not present at all times.

In the meantime, even applying a robust, standardized protocols that can generate a highly homogeneous cardiomyocyte population within a short period of time will not necessarily eliminate the intra-individual variation due to different hiPSC clones, differentiation batches, and even wells or surface used (plastic vs glass). This is also a widely recognized issue in the field of iPSC disease modeling, which greatly hinder the application of this technology especially when such application is of clinical relevance. In our study, we were fortunate to see similar EP results from the iPS-CMs (over 3 months old) from both groups regardless of how the data was combined, including data combined from CMs that derived from multiple iPSC clones (for S37P), from different batches, and from different wells. But nevertheless, it is too early to draw any definitive conclusion from our preliminary study without further repeating these experiments

with more iPSC lines added (especially for the control group). This is the main reason we carried out more functional characterizations using iPS-CMs derived directly from glass-bottom wells, and with the addition of two more WT control iPSC lines. In this yet-to-be-analyzed experiments, we also increased the characterization throughput to 96-well plates, aiming to further increase the statistical power of our data.

The use of isogenic controls for iPSC disease modeling studies

Besides the intra-individual variation mentioned above, a separate concern of disease modeling research especially when using iPSCs is the design of optimal control settings because of the inter-individual variation. Due to differences in genetic and epigenetic backgrounds, culture age *in vitro*, and ability to differentiate into the desired cell type, phenotypic differences observed between the patient's iPSC-derived cells and wild type control (individuals without the disease) iPSC-derived cells may not always be relevant to disease (Musunuru, 2013). Therefore, a more rigorous characterization of the functional consequence of the genetic alteration that contributes to the disease is to establish isogenic control cells that differ from the patients' cells only at the disease mutation, to maximally minimize the effect of the above potential confounders, and directly associate genotype with phenotype for disease causality investigation (Reinhardt et al., 2013). Traditional gene-targeting technologies based on homologous recombination proved to be technically challenging in iPSCs (Zwaka & Thomson, 2003). However, with the recently emerging high efficient CRISPR-Cas genome editing systems (Cho et al., 2013; Cong et al., 2013; Jinek et al., 2012; Jinek et al., 2013; Mali et al., 2013), the generation of such isogenic controls seems to become a more achievable goal (Hendriks et al., 2016).

Our lab (I was not involved in any of these projects) has spent over a year and half applying CRISPR-Cas based tools for creating NAA10 S37P knock-in models using mammalian cell lines without any success, validating that the efficiency of gene editing via CRISPR system can be target site (gene) specific, and to achieve base pair editing through template-based (e.g. single-stranded oligo DNA nucleotide) homology-directed repair (HDR) can be even more difficult (Cong et al., 2013; Cox et al., 2015; Hilton & Gersbach, 2015). An enhanced editing efficiency of point mutation correction with CRISPR-Cas is a must before isogenic control use in disease studies becomes routine, since majority of diseases being modeled now are still genetic diseases that arisen from point mutations in the genome. Some of the recent studies have reported progresses in this regard, providing hope for such applications as well as the potential therapeutic use of genome editing in the future (Komor et al., 2016; Paquet et al., 2016). Of course, one could also try to establish the isogenic control cells using other genome editing systems such as ones based on zinc finger proteins (ZNFs) or transcription activator-like effector nucleases (TALENs); however, it is important to recognize the limitations of these methods during application such as low-specificity and interference between contiguous modules in larger arrays (Chadwick & Musunuru, 2017; Gaj et al., 2013).

Proof-of-concept study and future directions

Our study discussed in this chapter reveals insights into the complicated genetics underlying Ogden syndrome especially from a cardiac standpoint; however, I want to emphasize that this work is still at its infancy, and has a "proof of concept" nature if considering all the limitations I discussed above. One should recognize that the data relate to our iPS-CM model are preliminary and need to be future investigated with additional evidences. Therefore, a careful interpretation of these results is highly recommended.

Besides the prominent cardiac dysfunction, Ogden syndrome patients also present significant neurological abnormalities, including neonatal hypotonia, poor feeding, apneic episodes, aspiration, cerebral atrophy, and microcephaly. One of the phenotype that I found particularly interesting is the brain MRI finding of the incomplete myelination of the splenium (the posterior portion of the corpus callosum) in one of these patients (the only medical record that is available to us), whereas the truncus (between splenium and genu) and genu (the anterior portion of the corpus callosum) show normal myelination. This brings to mind the consideration of dys- or demyelinating disorder since the splenium myelinates prior to myelination of the corpus callosum truncus and genu (Barkovich et al., 1988). Although the patient did suffer from respiratory distress (increased work of breathing) for a few weeks around that time (due to atelectasis, reactive airway diseases, aspiration, etc.), this specific demyelination pattern of his brain was less likely to be caused by hypoxia and ischemia.

Note that, the deficiency (abolished activity) of one of the proteins of interest (affected NatA substrates) mentioned in Chapter 3, aspartate-glutamate carrier 1 (AGC1), has been shown to be associated with a genetic disorder featuring arrested psychomotor development, epilepsy, hypotonia, and global hypomyelination in the cerebral hemispheres, many of which overlap with Ogden syndrome patients (Falk et al., 2014; Wibom et al., 2009; Wolf & van der Knaap, 2009). Therefore, it would be also valuable to establish the neuronal lineages differentiation from the S37P iPSCs model that we created for investigating the neurological phenotype of Ogden syndrome in-a-dish. A similar approach can be applied to examine other affected tissues in Ogden syndrome, such as adipocytes (fat tissue), since S37P patients have very little subcutaneous fat which partially gave them the “little old man” appearance; hematopoietic stem cells, so one can study the XCI-skewing phenomenon in Ogden syndrome carriers; smooth

muscle cells, to check whether S37P cells exhibit any premature senescence phenotypes like in the case of progeria diseases. Furthermore, to develop specific *in vitro* organoids from S37P iPSCs to model Ogden syndrome is another exciting future step one could explore, which will likely provide evidences that can better reflect the *in vivo* physiological and pathological situations (Clevers, 2016; Fatehullah et al., 2016; Hartley & Brennand, 2016; Muthuswamy, 2017).

Clinical insights

From a clinical point of view, as mentioned above, Ogden syndrome could be considered as a new type of long QT syndrome, and also a severe form. Like many LQTS patients, OS patients also present the dynamic phenotypic expression and heterogeneous process of ventricular repolarization, which explains 14/17 ECGs of the OS patient I analyzed showing normal QT intervals, which might also be the reason why not all OS patients had documented prolonged QT intervals in their medical record. For such patients, a single recorded QTc value within a normal range for sure does not rule out the risks of developing a long QTc and malignant cardiac arrhythmia later in life. In fact, study has shown that the maximum QTc duration measured at any time before age 10 of a long QT syndrome patient is the most powerful predictor of cardiac events during adolescence, regardless of baseline, mean, or most recent QTc values (Goldenberg et al., 2006). Hence, we recommend all patients with NAA10 variants to be continually monitored and assessed via ECGs especially before 10 years of age, and avoid risk factors that will prolong the QT interval such as certain medications, which might potentially improve their prognosis.

4.5 Concluding Remarks

In conclusion, I report here the creation and initial characterization of an iPS-CM model for Ogden syndrome and introduced the first application of OptoDyCE platform in investigating patient specific iPSC-derived cardiomyocytes. We demonstrated the defects of S37P-H iPSCs in forming embryoid bodies in suspended culture and initiating cell differentiation into the three embryonic germ layers *in vitro*, and the enrichment of favorable substrates of NatA during reprogramming. I recommended the recognition of Ogden syndrome as a new type of long QT syndrome, hence these patients should be managed and treated accordingly if future families are identified. We provided additional evidence of the high variability of patient's QT intervals *in vivo* and the electrical instability of the patient's iPS-CMs *in vitro*. Our proof-of-concept study of the S37P iPS-CM model successfully recapitulated the cardiac phenotype of the patients at the cellular level including the QT prolongation at a longer cycle length, suggesting an imbalance of depolarization and hyperpolarizing currents in patient's cells. We hypothesize that the cardiac phenotype of Ogden syndrome is caused by misregulation of key ion channel proteins under the effect of a disrupted Nt-acetylation pathway.

Mechanism wise, both iPSCs and iPS-CMs derived from the Ogden syndrome patient can serve as great platforms for continually addressing many of the outstanding questions. Future work could include establishing the isogenic control cells, performing additional analysis on other ion channels especially the calcium handling pathway, investigating the metabolic pathways including bioenergetics profiling, introducing WT NAA10 to S37P iPS-CMs to rescue the phenotype, investigating the knockdown effect of NAA10 using shRNA or siRNA (and compare the phenotype with S37P derived cells), *in vitro* proteome-wide or targeted protein Nt-acetylation screening of NatA substrates in patient's derived cells, and evaluating the interaction

and regulation between substrates of interest and NAA10 by routine biochemical assays. Moreover, one can also investigate the role that NAA10 and/or the S37P mutant plays in other major disease-affected tissues such as neuronal cells by differentiating our S37P iPSCs into specific cell lineages of interest or even organoids. Last but not least, NAA10 S37P knock-in and NAA10 knockout mice models could be explored in parallel as well to gather additional evidences for helping understand Ogden syndrome pathogenesis at a systematic level.

Chapter 5

Conclusions

In this doctoral dissertation, I pursued four seemingly distinct but intimately connected studies aimed at assessing and improving some of the challenges of implementing precision medicine. The first study focused on developing a high throughput targeted resequencing protocol to help produce a set of true variants to facilitate research discovery and clinical decision making. The second study focused on implementing the developed strategy of generating reliable NGS analysis results to accelerate the discovery of the genetic basis of diseases. The third study focused on demonstrating the downstream functional investigation of a disease-contributory variant identified through clinical genomics. And the final study focused on

creating a personalized disease model to facilitate the understanding of disease mechanisms and the design of optimal therapies in the future.

In chapter 1, we found despite the popular use of high throughput DNA sequencing technologies in both research and clinical settings, there are still many uncertainties related to the quality of NGS data due to sequencing biases and algorithm artifacts. Our successfully developed high-throughput validation protocol enabled us to evaluate the reliability and comparability between different sequencing platforms and variant calling tools. Our analysis revealed widely discordant variants generated by varying platforms and computational tools when applied to the same sample or data set, and the situation of Indel calls are worse than SNVs. Moreover, when compared WGS and WES in calling Indels from the same data set, the resulting variants have a high discordance rate. And in general, we found that the accuracy of INDEL detection with WGS is much better than WES. Based on these observations, we offered a few practical suggestions to improve the data quality under the current technologies, which includes to use a similar validation protocol as ours to evaluate variant sets, select WGS to produce sequences for Indel studies, and to use multiple analytic pipelines to reduce errors in variant calling.

In chapter 2, we first introduced the discovery of a new X-linked disorder involving TAF1 mutations via clinical genomics. Patients with this syndrome have characteristic features including global developmental delay, intellectual disability, characteristic facial dysmorphologies, generalized hypotonia, and peculiar gluteal creases with sacral caudal remnants. We highlighted the benefit of using large pedigrees to reliably eliminate false variant calls, and of sharing detailed clinical case reports on a preprint server in an effort to find more patients and to set up collaborations. In the second project, we highlighted the combined use of

WGS and HPO to maximize the chance of locating disease relevant variants. However, at the same time, we also revealed the limitations of using clinical genomics to pinpoint a causative variant, especially when the phenotype is ambiguous and reliable clinical data is far from reach for optimized variant prioritization. Lastly, we recommended that researchers interpret and report positive clinical genomics findings with extra caution as mistakes in the reference sequence can occur.

In chapter 3, I reported a detailed phenotyping analysis of Ogden syndrome, and a genotype-phenotype correlation analysis of NAA10-related disorders. We produced evidence illustrating Ogden syndrome is linked to downstream N-terminal acetylation defects. To be specific, we found that NAA10 S37P mediates a less efficient interaction with its auxiliary partner NAA15 in forming NatA complex when compared to the WT, which likely leads to the overall reduction of its enzymatic activity. Moreover, NAA10 S37P fibroblasts show reduced growth and proliferation under routine culture. And when cultured in different densities, S37P cells altered the metabolic activity in an unusual way compared to the WT controls, which might be a stress-related phenotype. Lastly, I revealed that S37P fibroblasts have an elevated glycolysis level, despite having a normal mitochondrial respiration at the same time. I proposed to further characterize the glucose metabolism pathway in S37P cells in order to determine the specific cause that contributes to the glycolysis-primed phenotype in these cells. Besides, I also discussed the necessity of performing rescue and knockdown experiments to validate the phenotype we saw in S37P cells is disease-specific. This is especially crucial for our study because we are aware that only a limited number of patient and (non-isogenic) control cell lines were used in most of our experiments. Moreover, additional biochemistry characterization of specific protein targets will also be useful to reveal the molecular path to the disease phenotype.

In chapter 4, we introduced the creation of patient specific iPS-derived cardiomyocytes model for investigating the cardiac phenotype of Ogden syndrome in-a-dish. I first demonstrated Ogden syndrome patients have high QT variability, which is a classic feature of congenital long QT syndrome (LQTS). And I suggested that Ogden syndrome be considered as a new subtype of LQTS due to misregulations of major cardiac channel proteins under a disrupted Nt-acetylation pathway. We next employed OptoDyCE platform for characterizing the electrophysiological properties of Ogden syndrome patient-derived cardiomyocytes, which we highlighted is the first application of OptoDyCE in studying patient specific iPS-CMs. Our initial electrophysiological experiments suggest prolonged action potentials in patient iPS-CMs that only unmasked at slow pacing rates, and a preserved morphology of calcium transients in patient iPS-CMs but with slightly slower upstroke. The consequences of these findings are under study, and we hope new results (especially from the sodium channel inhibition experiment) will provide us further insights into the mechanism. The developed iPS-CMs model of Ogden syndrome and the employed all-optical platform OptoDyCE permit the first mechanistic investigation of the effects of N-terminal acetylation on cardiac electrophysiology and arrhythmias. And I also hypothesized a temporal and spatial regulation of Nt-acetylation modification is a possibility during development and disease pathogenesis in general, which can be further explored using S37P iPSCc derivatives that belong to other disease-relevant organs and tissues.

The work discussed in chapter 3 and 4 revealed several common challenges of disease modeling studies. First is the limited statistical power of data generated from a very small patient and control cohorts, the latter usually is contributed by factors such as the rare nature of the disease studied, the difficulty of collecting relevant samples, limited funding and manpower. Second is the lack of genetically matched, isogenic controls used in most studies. Both of the

above two things reduce the confidence of identifying true pathological changes in the patients' cells since genetic background differences and line-to-line variations can both contribute to phenotypic differences to some degree and confound the findings. This can be especially problematic when disease-relevant differences are anticipated to be subtle (Hockemeyer & Jaenisch, 2016). This situation can be changed with the use of rapidly developing genome editing technologies, such as ZFN, TALENs, and CRISPR-Cas9 systems. Using these approaches, one can introduce the disease-contributory mutations into non-disease affected WT cells, and also correct the specific genetic changes in patient-derived cells. The latter also holds the promise for gene therapy, although many technical challenges, such as off-target effects, still need to be overcome before that can be readily used in the clinic (Cornu et al., 2017; Cox et al. 2015; Hotta & Yamanaka, 2015; Prakash et al., 2016).

In chapter 4, I also discussed two major issues concerning the current hiPSC research: the immaturity and heterogeneity of hiPSC derivatives, both of which are obstacles for hiPSCs achieving their full potential in advancing the practice of precision medicine (Avior et al., 2016; Chen et al., 2016; Sayed et al., 2016). For example, the embryonic nature of hiPSCs and fetal-like properties of hiPSC derivatives pose great challenges in modeling late-onset disorders and age-associated phenotypes (Studer et al., 2015), such as Parkinson's disease and late-onset Alzheimer's. Also, the combination of unpredictable clonal variation of hiPSC lines and non-standardized differentiation protocols used oftentimes result in a heterogeneous population of targeted cells (as discussed in chapter 4), which will cause considerable problems during high-content drug screening where large-scale of cells presenting a consistent and robust phenotype is the key to success (Hung et al., 2017; Sayed et al., 2016; Shi et al., 2016; Suh, 2017).

During the course of time of this dissertation work (2012.6 – 2017.4), we continue to witness the advancement of precision medicine from bench to bedside. New technologies are rapidly becoming available to propel precision medicine forward: e.g. the advances of blood-based liquid biopsies enable rapid and accurate detection of cancer mutations and metastasis (Siravegna et al., 2017; Wan et al., 2017); artificial intelligence and machine learning algorithms (e.g. IBM's Watson) have been employed for reliably and efficiently analyzing and interpreting health data (Rhrissorrakrai et al., 2016); the world's first quantum computing machine was born as this dissertation revision was just finished (Wang et al., 2017), which might revolutionize the way health-related big data is stored and processed in the future. Moreover, a PMI-related national effort "Cancer Moonshot" was launched in January 2016, aiming to accelerate cancer research, prevention, treatment, and cure through precision oncology (Office of the Press Secretary, 2016). Although there are still substantial improvements must be made to achieve more effective precision medicine, Dr. Garrod's visionary of treating each patient uniquely "with the least possible damage" may soon become a reality.

References

Abrams, D. J., & Macrae, C. a. (2014). Long QT syndrome. *Circulation*, *129*(14), 1524–9.

Abriel, H., & Zaklyazminskaya, E. V. (2013). Cardiac channelopathies: genetic and molecular mechanisms. *Gene*, *517*(1), 1–11.

Abu-Amero, K. K., Hellani, A., Salih, M. A., Al Hussain, A., al Obailan, M., Zidan, G., … Bosley, T. M. (2010). Ophthalmologic abnormalities in a de novo terminal 6q deletion. *Ophthalmic Genet*, *31*(1), 1–11.

Aksnes, H., Drazic, A., Marie, M., & Arnesen, T. (2016). First things first: vital protein marks by N-terminal acetyltransferases. *Trends in Biochemical Sciences*, *41*(9), 746–760.

Aksnes, H., Van Damme, P., Goris, M., Starheim, K. K., Marie, M., Støve, S. I., … Arnesen, T. (2015). An organellar nα-acetyltransferase, naa60, acetylates cytosolic N termini of transmembrane proteins and maintains golgi integrity. *Cell Reports*, *10*(8), 1362–1374.

Allanson, J. E., Biesecker, L. G., Carey, J. C., & Hennekam, R. C. M. (2009). Elements of morphology: introduction. *American Journal of Medical Genetics. Part A*, *149A*(1), 2–5.

Allen, K. J., Gurrin, L. C., Constantine, C. C., Osborne, N. J., Delatycki, M. B., Nicoll, A. J., … Gertig, D. M. (2008). Iron-overload-related disease in HFE hereditary hemochromatosis. *New England Journal of Medicine*, *358*(3), 221–230.

Altman, R. B., Prabhu, S., Sidow, A., Zook, J. M., Goldfeder, R., Litwack, D., … Kass-Hout, T. (2016). A research roadmap for next-generation sequencing informatics. *Science Translational Medicine*, *8*(335), 10–335.

Ambrosi, C. M., & Entcheva, E. (2014). Optogenetic control of cardiomyocytes via viral delivery. *Methods in Molecular Biology*, (1181), 215–228.

Amendola, L. M., Jarvik, G. P., Leo, M. C., McLaughlin, H. M., Akkari, Y., Amaral, M. D., … Rehm, H. L. (2016). Performance of ACMG-AMP variant-interpretation guidelines among

nine laboratories in the clinical sequencing exploratory research consortium. *American Journal of Human Genetics*, *98*(6), 1067–76.

Apostolou, E., & Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. *Nature*, *502*(7472), 462–71.

Arnar, D. O., & Palsson, R. (2017). Genetics of common complex diseases: a view from Iceland. *European Journal of Internal Medicine*, pii: S0953-6205(17)30122-X. [Epub ahead of print].

Arnaudo, N., Fernández, I. S., McLaughlin, S. H., Peak-Chew, S. Y., Rhodes, D., & Martino, F. (2013). The N-terminal acetylation of Sir3 stabilizes its binding to the nucleosome core particle. *Nature Structural & Molecular Biology*, *20*(9), 1119–21.

Arnesen, T., Anderson, D., Torsvik, J., Halseth, H. B., Varhaug, J. E., & Lillehaug, J. R. (2006). Cloning and characterization of hNAT5/hSAN: an evolutionarily conserved component of the NatA protein N-alpha-acetyltransferase complex. *Gene*, *371*(2), 291–5.

Arnesen, T., Starheim, K. K., Van Damme, P., Evjenth, R., Dinh, H., Betts, M. J., … Anderson, D. (2010). The chaperone-like protein HYPK acts together with NatA in cotranslational N-terminal acetylation and prevention of Huntingtin aggregation. *Molecular and Cellular Biology*, *30*(8), 1898–909.

Arnesen, T., Van Damme, P., Polevoda, B., Helsens, K., Evjenth, R., Colaert, N., … Gevaert, K. (2009). Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(20), 8157–62.

Ashley, E. A. (2016). Towards precision medicine. *Nature Reviews. Genetics*, *17*(9), 507–22.

Atmanli, A., & Domian, I. J. (2016). Recreating the cardiac microenvironment in pluripotent stem cell models of human physiology and disease. *Trends in Cell Biology*, *31*(0), 3613–3621.

Avior, Y., Sagi, I., & Benvenisty, N. (2016). Pluripotent stem cells in disease modelling and drug discovery. *Nature Reviews Molecular Cell Biology*, *17*(3), 170–182.

Bader, P. L., Faizi, M., Kim, L. H., Owen, S. F., Tadross, M. R., Alfa, R. W., … Shamloo, M. (2011). Mouse model of Timothy syndrome recapitulates triad of autistic traits. *Proceedings of the National Academy of Sciences*, *108*(37), 15432–15437.

Baker, M. W., Atkins, A. E., Cordovado, S. K., Hendrix, M., Earley, M. C., & Farrell, P. M. (2016). Improving newborn screening for cystic fibrosis using next-generation sequencing technology: a technical feasibility study. *Genetics in Medicine*, *18*(3), 231–8.

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A., & Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*, *12*(11), 745–755.

Barkovich, A. J., Kjos, B. O., Jackson, D. E., & Norman, D. (1988). Normal maturation of the

neonatal and infant brain: MR imaging at 1.5 T. *Radiology*, *166*(1), 173–80.

Basho, R. K., Eterovic, A. K., & FundaMeric-Bernstam. (2015). Clinical applications and limitations of next-generation sequencing. *The American Journal of Hematology/Oncology*, *11*(3), 17–22.

Bass, B. G. (1975). Restitution of the action potential in cat papillary muscle. *The American Journal of Physiology*, *228*(6), 1717–1724.

Bayart, E., & Cohen-Haguenauer, O. (2013). Technological overview of iPS induction from human adult somatic cells. *Current Gene Therapy*, *13*(2), 73–92.

Bazett, H. (1920). An analysis of the time-relations of electrocardiograms. *Heart*, *7*, 353–370.

Behnia, R., Panic, B., Whyte, J. R. C., & Munro, S. (2004). Targeting of the Arf-like GTPase Arl3p to the Golgi requires N-terminal acetylation and the membrane protein Sys1p. *Nature Cell Biology*, *6*(5), 405–13.

Bellin, M., Marchetto, M. C., Gage, F. H., & Mummery, C. L. (2012). Induced pluripotent stem cells: the new patient? *Nature Reviews Molecular Cell Biology*, *13*(11), 713–726.

Beltran-Alvarez, P., Tarradas, A., Chiva, C., P??rez-Serra, A., Batlle, M., P??rez-Villa, F., … Pagans, S. (2014). Identification of N-terminal protein acetylation and arginine methylation of the voltage-gated sodium channel in end-stage heart failure human heart. *Current Therapeutic Research - Clinical and Experimental*, *76*, 126–129.

Ben-Ari, M., Naor, S., Zeevi-Levin, N., Schick, R., Jehuda, R. Ben, Reiter, I., … Binah, O. (2016). Developmental changes in electrophysiological characteristics of human induced Pluripotent Stem Cell-derived cardiomyocytes. *Heart Rhythm Society*, *13*(12), 2379–2387.

Benjamin, E. J., Blaha, M. J., Chiuve, S. E., Cushman, M., & Das, S. R. et al. (2017). *Heart disease and stroke statistics--2017 update: a report from the American Heart Association*. *Circulation* (Vol. 135).

Berridge, M. V, Herst, P. M., & Tan, A. S. (2005). Tetrazolium dyes as tools in cell biology: new insights into their cellular reduction. *Biotechnology Annual Review*, *11*, 127–52.

Bers, D. M. (2002). Cardiac excitation-contraction coupling. *Nature*, *415*(6868), 198–205.

Bezzerides, V. J., Zhang, D., Pu, W. T., Blake, R., Park, J., & Trayanova, N. (2017). Modeling inherited arrhythmia disorders using induced pluripotent stem cell-derived cardiomyocytes. *Circulation Journal*, *81*(1), 12–21.

Biesecker, L. G. (2005). Mapping phenotypes to language: a proposal to organize and standardize the clinical descriptions of malformations. *Clinical Genetics*, *68*(4), 320–6.

Biesecker, L. G., & Green, R. C. (2014). Diagnostic clinical genome and exome sequencing. *New England Journal of Medicine*, *370*, 2418–2425.

Blazeski, A., Zhu, R., Hunter, D. W., Weinberg, S. H., Boheler, K. R., Zambidis, E. T., & Tung, L. (2012). Electrophysiological and contractile function of cardiomyocytes derived from

human embryonic stem cells. *Progress in Biophysics and Molecular Biology*, *110*(2–3), 178–195.

Blumenthal, G. M., Mansfield, E., & Pazdur, R. (2016). Next-generation sequencing in oncology in the era of precision medicine. *JAMA Oncology*, *2*(1), 13–4.

Bodian, D. L., Klein, E., Iyer, R. K., Wong, W. S. W., Kothiyal, P., Stauffer, D., … Solomon, B. D. (2016). Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates. *Genetics in Medicine*, *18*(3), 221–30.

Bohnen, M. S., Peng, G., Robey, S. H., Terrenoire, C., Iyer, V., Sampson, K. J., & Kass, R. S. (2017). Molecular pathophysiology of congenital long QT syndrome. *Physiological Reviews*, *97*(1), 89–134.

Bolze, A., Byun, M., McDonald, D., Morgan, N. V, Abhyankar, A., Premkumar, L., … Casanova, J.-L. (2010). Whole-exome-sequencing-based discovery of human FADD deficiency. *American Journal of Human Genetics*, *87*(6), 873–81.

Bolze, A., Mahlaoui, N., Byun, M., Turner, B., Trede, N., Ellis, S. R., … Casanova, J.-L. (2013). ribosomal protein SA haploinsufficiency in humans with isolated congenital asplenia. *Science*, *340*(6135), 976–978.

Boycott, K. M., Vanstone, M. R., Bulman, D. E., & MacKenzie, A. E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet*, *14*(10), 681–691.

Boyd, S. D., Galli, S. J., Schrijver, I., Zehnder, J. L., Ashley, E. A., & Merker, J. D. (2014). A balanced look at the implications of genomic (and other "Omics") testing for disease diagnosis and clinical care. *Genes*, *5*(3), 748–66.

Brand, M. D. (1990). The proton leak across the mitochondrial inner membrane. *Biochimica et Biophysica Acta*, *1018*(2–3), 128–133.

Burridge, P. W., Keller, G., Gold, J. D., & Wu, J. C. (2012). Production of de novo cardiomyocytes: human pluripotent stem cell differentiation and direct reprogramming. *Cell Stem Cell*, *10*(1), 16–28.

Burridge, P. W., Matsa, E., Shukla, P., Lin, Z. C., Churko, J. M., Ebert, A. D., … Wu, J. C. (2014). Chemically defined generation of human cardiomyocytes. *Nature Methods*, *11*(8), 855–860.

Burton, R. A. B., Klimas, A., Ambrosi, C. M., Tomek, J., Corbett, A., Entcheva, E., & Bub, G. (2015). Optical control of excitation waves in cardiac tissue. *Nature Photonics*, *9*(12), 813–816.

Byun, M., Abhyankar, A., Lelarge, V., Plancoulaine, S., Palanduz, A., Telhan, L., … Casanova, J.-L. (2010). Whole-exome sequencing-based discovery of STIM1 deficiency in a child with fatal classic Kaposi sarcoma. *The Journal of Experimental Medicine*, *207*(11), 2307–12.

Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., … Lander, E. S. (1999). Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet*, *22*(3), 231–238.

Carnevali, P., Baccash, J., Halpern, A. L., Nazarenko, I., Nilsen, G. B., Pant, K. P., … Drmanac, R. (2012). Computational techniques for human genome resequencing using mated gapped reads. *Journal of Computational Biology*, *19*(3), 279–292.

Carrel, L., & Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females, *434*(March).

Casey, J. P., Støve, S. I., McGorrian, C., Galvin, J., Blenski, M., Dunne, A., … Lynch, S. A. (2015). NAA10 mutation causing a novel intellectual disability syndrome with Long QT due to N-terminal acetyltransferase impairment. *Scientific Reports*, *5*, 16022.

Caspi, O., Huber, I., Gepstein, A., Arbel, G., Maizels, L., Boulos, M., & Gepstein, L. (2013). Modeling of arrhythmogenic right ventricular cardiomyopathy with human induced pluripotent stem cells. *Circulation. Cardiovascular Genetics*, *6*(6), 557–68.

Cassidy, S. B., Schwartz, S., Miller, J. L., & Driscoll, D. J. (2012). Prader-Willi syndrome. *Genet Med*, *14*(1), 10–26.

Chadwick, A. C., & Musunuru, K. (2017). Genome editing for the study of cardiovascular diseases. *Current Cardiology Reports*, *19*(3), 22.

Char, D. S. (2015). Whole-genome sequencing in critically ill infants and emerging ethical challenges. *The Lancet Respiratory Medicine*.

Chen, I. Y., Matsa, E., & Wu, J. C. (2016). Induced pluripotent stem cells: at the heart of cardiovascular precision medicine. *Nature Reviews Cardiology*, *13*(6), 333–349.

Chen, V., Staub, R. E., Fong, S., Tagliaferri, M., Cohen, I., & Shtivelman, E. (2012). Bezielle selectively targets mitochondria of cancer cells to inhibit glycolysis and OXPHOS. *PLoS ONE*, *7*(2), 1–12.

Cheng, E. P., Yuan, C., Navedo, M. F., Dixon, R. E., Nieves-Cintron, M., Scott, J. D., & Santana, L. F. (2011). Restoration of normal L-type Ca^{2+} channel function during Timothy syndrome by ablation of an anchoring protein. *Circulation Research*, *109*(3), 255–261.

Chinwalla, A. T., Cook, L. L., Delehaunty, K. D., Fewell, G. A., Fulton, L. A., Fulton, R. S., … Zody, M. C. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, *420*(6915), 520–562.

Cho, S. W., Kim, S., Kim, J. M., & Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnology*, *31*(3), 230–2.

Christian, S. L., Robinson, W. P., Huang, B., Mutirangura, A., Line, M. R., Nakao, M., … Ledbetter, D. H. (1995). Molecular characterization of two proximal deletion breakpoint regions in both Prader-Willi and Angelman syndrome patients. *American Journal of Human Genetics*, *57*(1), 40–48.

Cirulli, E. T., & Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nature Reviews Genetics*, *11*(6), 415–425.

Clancy, C. E., & Rudy, Y. (1999). Linking a genetic defect to its cellular phenotype in a cardiac arrhythmia. *Nature*, *400*(6744), 566–569.

Clancy, C. E., Tateyama, M., & Kass, R. S. (2002). Insights into the molecular mechanisms of bradycardia-triggered arrhythmias in long QT-3 syndrome. *Journal of Clinical Investigation*, *110*(9), 1251–1262.

Clark, M. J., Chen, R., Lam, H. Y., Karczewski, K. J., Euskirchen, G., Butte, A. J., & Snyder, M. (2011). Performance comparison of exome DNA sequencing technologies. *Nature Biotechnology*, *29*(10), 908–914.

Clement, N. L., Snell, Q., Clement, M. J., Hollenhorst, P. C., Purwar, J., Graves, B. J., … Johnson, W. E. (2009). The GNUMAP algorithm: unbiased probabilistic mapping of oligonucleotides from next-generation sequencing. *Bioinformatics*, *26*(1), 38–45.

Clevers, H. (2016). Modeling development and disease with organoids. *Cell*, *165*(7), 1586–1597.

Collins, F. S. (2014). Francis Collins says medicine in the future will be tailored to your genes. The Wall Street Journal. Retrieved July 7, 2017, from https://www.wsj.com/articles/francis-collins-says-medicine-in-the-future-will-be-tailored-to-your-genes-1404763139.

Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: lessons from large-scale biology. *Science*, *300*(5617), 286–290.

Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. *New England Journal of Medicine*, *372*(9), 793–795.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., … Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819–823.

Corbett-Detig, R. B., Zhou, J., Clark, A. G., Hartl, D. L., & Ayroles, J. F. (2013). Genetic incompatibilities are widespread within species. *Nature*, *504*(7478), 135–137.

Cornish, A., & Guda, C. (2015). A comparison of variant calling pipelines using genome in a bottle as a reference. *BioMed Research International*, *2015*, 456479.

Cornu, T. I., Mussolino, C., & Cathomen, T. (2017). Refining strategies to translate genome editing to the clinic. *Nature Medicine*, *23*(4), 415–423.

Cox, D. B. T., Platt, R. J., & Zhang, F. (2015). Therapeutic genome editing: prospects and challenges. *Nature Medicine*, *21*(2), 121–131.

Davis, R. P., Casini, S., van den Berg, C. W., Hoekstra, M., Remme, C. A., Dambrot, C., … Mummery, C. L. (2012). Cardiomyocytes derived from pluripotent stem cells recapitulate electrophysiological characteristics of an overlap syndrome of cardiac sodium channel disease. *Circulation*, *125*(25), 3079–3091.

Delaney, S. K., Hultner, M. L., Jacob, H. J., Ledbetter, D. H., McCarthy, J. J., Ball, M., … Green,

R. C. (2016). Toward clinical genomics in everyday medicine: perspectives and recommendations. *Expert Review of Molecular Diagnostics*, *16*(5), 521–532.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–8.

Dickmann, L. J., & Ware, J. A. (2016). Pharmacogenomics in the age of personalized medicine. *Drug Discovery Today: Technologies*, *21–22*, 11–16.

Dietrich, P., & Dragatsis, I. (2016). Familial dysautonomia: mechanisms and models. *Genetics and Molecular Biology*, *39*(4), 497–514.

Dikiy, I., & Eliezer, D. (2014). N-terminal acetylation stabilizes N-terminal helicity in lipid- and micelle-bound α-synuclein and increases its affinity for physiological membranes. *The Journal of Biological Chemistry*, *289*(6), 3652–65.

Dong, L., Wang, W., Li, A., Kansal, R., Chen, Y., Chen, H., & Li, X. (2015). Clinical next generation sequencing for precision medicine in cancer. *Current Genomics*, *16*(4), 253–63.

Dörfel, M. J., & Lyon, G. J. (2015). The biological functions of Naa10 — From amino-terminal acetylation to human disease. *Gene*, *567*(2), 103–131.

Dranka, B. P., Hill, B. G., & Darley-Usmar, V. M. (2010). Mitochondrial reserve capacity in endothelial cells: the impact of nitric oxide and reactive oxygen species. *Free Radical Biology & Medicine*, *48*(7), 905–14.

Drazic, A., Myklebust, L. M., Ree, R., & Arnesen, T. (2016). The world of protein acetylation. *Biochimica et Biophysica Acta - Proteins and Proteomics*, *1864*(10), 1372–1401.

Du, D. T. M., Hellen, N., Kane, C., & Terracciano, C. M. N. (2015). Action potential morphology of human induced pluripotent stem cell-derived cardiomyocytes does not predict cardiac chamber specificity and is dependent on cell density. *Biophysical Journal*, *108*(1), 1–4.

Du, W., & Pogoriler, J. (2006). Retinoblastoma family genes. *Oncogene*, *25*(38), 5190–200.

Eble, T. N., Sutton, V. R., Sangi-Haghpeykar, H., Wang, X., Jin, W., Lewis, R. A., … Van den Veyver, I. B. (2009). Non-random X chromosome inactivation in Aicardi syndrome. *Human Genetics*, *125*(2), 211–6.

Edwin, C. (2016). Precision medicine: technology, regulations and challenges. Retrieved April 11, 2017, from http://www.raps.org/Regulatory-Focus/News/2016/03/11/24528/Precision-Medicine-Technology-Regulations-and-Challenges/.

Eisenberger, T., Neuhaus, C., Khan, A. O., Decker, C., Preising, M. N., Friedburg, C., … Bolz, H. J. (2013). Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies. *PloS One*, *8*(11), e78496.

El-Hattab, A. W., & Scaglia, F. (2016). Mitochondrial cardiomyopathies. *Frontiers in Cardiovascular Medicine*, *3*, 25.

Elharrar, V., & Surawicz, B. (1983). Cycle length effect on restitution of action potential duration in dog cardiac fibers. *The American Journal of Physiology*, *244*(6), H782-92.

Esmailpour, T., Riazifar, H., Liu, L., Donkervoort, S., Huang, V. H., Madaan, S., … Huang, T. (2014). A splice donor mutation in NAA10 results in the dysregulation of the retinoic acid signalling pathway and causes Lenz microphthalmia syndrome. *Journal of Medical Genetics*, *51*(3), 185–96.

Evjenth, R., Hole, K., Karlsen, O. a, Ziegler, M., Arnesen, T., & Lillehaug, J. R. (2009). Human Naa50p (Nat5/San) displays both protein N alpha- and N epsilon-acetyltransferase activity. *The Journal of Biological Chemistry*, *284*(45), 31122–9.

Falk, M. J., Li, D., Gai, X., Mccormick, E., Place, E., Lasorsa, F. M., … Hakonarson, H. (2014). AGC1 deficiency causes infantile epilepsy, abnormal myelination, and reduced N-acetylaspartate. *JIMD Reports*, *14*, 77–85.

Fang, H., Bergmann, E. A., Arora, K., Vacic, V., Zody, M. C., Iossifov, I., … Narzisi, G. (2016). Indel variant analysis of short-read sequencing data with Scalpel. *Nat Protoc*, *11*(12), 2529–2548.

Fang, H., Wu, Y., Narzisi, G., O'Rawe, J. A., Barrón, L. T. J., Rosenbaum, J., … Lyon, G. J. (2014). Reducing INDEL calling errors in whole-genome and exome sequencing data. *Genome Medicine*, *6*(6), 89.

Fang, H., Wu, Y., Yang, H., Yoon, M., Jiménez-Barrón, L. T., Mittelman, D., … Lyon, G. J. (2017). Whole genome sequencing of one complex pedigree illustrates challenges with genomic medicine. *BMC Medical Genomics*, *10*(1), 10.

Fatehullah, A., Tan, S. H., & Barker, N. (2016). Organoids as an *in vitro* model of human development and disease. *Nature Cell Biology*, *18*(3), 246–54.

Fatima, A., Xu, G., Shao, K., Papadopoulos, S., Lehmann, M., Arnáiz-Cot, J. J., … Saric, T. (2011). *In vitro* modeling of ryanodine receptor 2 dysfunction using human induced pluripotent stem cells. *Cellular Physiology and Biochemistry*, *28*(4), 579–92.

Federal Policy for the Protection of Human Subjects ('Common Rule'). (1991). Retrieved April 11, 2017, from https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html.

Fernández-Cañón, J. M., Granadino, B., De Bernabé, D. B., Renedo, M., Fernández-Ruiz, E., Peñalva, M. A., & De Córdoba, S. R. (1996). The molecular basis of alkaptonuria. *Nature Genetics*, *14*(1), 19–24.

Fisher, T. S., Etages, S. Des, Hayes, L., Crimin, K., & Li, B. (2005). Analysis of ARD1 function in hypoxia response using retroviral RNA interference. *The Journal of Biological Chemistry*, *280*(18), 17749–57.

Food and Drug Administration, U. D. of H. and H. S. (2013). Paving the way for personalized medicine: FDA's role in a new era of medical product development.

FRANZ, M. R. (2003). The electrical restitution curve revisited: steep or flat slope-which is better? *Journal of Cardiovascular Electrophysiology*, *14*(s10), S140–S147.

Freund, J., Brandmaier, A. M., Lewejohann, L., Kirste, I., Kritzler, M., Krüger, A., … Kempermann, G. (2013). Emergence of individuality in genetically identical mice. *Science*, *340*(6133), 756–759.

Funderburk, S. F., Wang, Q. J., & Yue, Z. (2010). The Beclin 1-VPS34 complex--at the crossroads of autophagy and beyond. *Trends in Cell Biology*, *20*(6), 355–62.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., & Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, *85*(8), 348–62.

Futema, M., Plagnol, V., Whittall, R. A., Neil, H. A. W., Humphries, S. E., Humphries, S. E., & UK10K. (2012). Use of targeted exome sequencing as a diagnostic tool for Familial Hypercholesterolaemia. *Journal of Medical Genetics*, *49*(10), 644–649.

Gagan, J., & Van Allen, E. M. (2015). Next-generation sequencing to guide cancer therapy. *Genome Medicine*, *7*(1), 80.

Gage, F. (2010). The promise and the challenge of modelling human disease in a dish. *EMBO Molecular Medicine*, *2*(3), 77–8.

Gaj, T., Gersbach, C. A., Barbas, C. F., & III. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, *31*(7), 397–405.

Gargis, A. S., Kalman, L., & Lubin, I. M. (2016). Assuring the quality of next-generation sequencing in clinical microbiology and public health laboratories. *Journal of Clinical Microbiology*, *54*(12), 2857–2865.

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv Preprint arXiv:1207.3907*.

Garrod, A. (1902). The incidence of alkaptonuria: a study in chemical individuality. *Lancet*, *160*(4137), 1616–1620.

Garrod, A. E. (1909). *Inborn Errors of Metabolism*. Henry Frowde and Hodder & Stoughton: London.

Garrod, A. E. (1931). *The inborn factors in disease: an essay.* Oxford University Press: Oxford, UK.

Gaudesius, G., Miragoli, M., Thomas, S. P., & Rohr, S. (2003). Coupling of cardiac electrical activity over extended distances by fibroblasts of cardiac origin. *Circulation Research*, *93*(5), 421–428.

Gautschi, M., Just, S., Mun, A., Ross, S., Rücknagel, P., Dubaquié, Y., … Dubaquie, Y. (2003). The yeast N(alpha)-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides. *Molecular and Cellular Biology*, *23*(20), 7403–14.

Gelb, B. D., & Tartaglia, M. (2006). Noonan syndrome and related disorders: dysregulated RAS-mitogen activated protein kinase signal transduction. *Human Molecular Genetics*, *15*(Review Issue 2), R220-6.

Gelzer, A. R. M., Koller, M. L., Otani, N. F., Fox, J. J., Enyeart, M. W., Hooker, G. J., … Gilmour, R. F. (2008). Dynamic mechanism for initiation of ventricular fibrillation *in vivo*. *Circulation*, *118*(11), 1123–1129.

Genome of the Netherlands Consortium, L. C., Menelaou, A., Pulit, S. L., van Dijk, F., Palamara, P. F., Elbers, C. C., … Wijmenga, C. (2014). Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nature Genetics*, *46*(8), 818–25.

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W. M., Willemsen, M. H., … Veltman, J. A. (2014). Genome sequencing identifies major causes of severe intellectual disability. *Nature*, *511*(7509), 344–7.

Gimm, O., Greco, A., Hoang-Vu, C., Dralle, H., Pierotti, M. A., & Eng, C. (1999). Mutation analysis reveals novel sequence variants in NTRK1 in sporadic human medullary thyroid carcinoma. *The Journal of Clinical Endocrinology & Metabolism*, *84*(8), 2784–2787.

Giudicessi, J., & Ackerman, M. (2013). Genetic testing in heritable cardiac arrhythmia syndromes: differentiating pathogenic mutations from background genetic noise. *Current Opinion in Cardiology*, *28*(1), 63–71.

Goedel, A., My, I., Sinnecker, D., & Moretti, A. (2017). Perspectives and challenges of pluripotent stem cells in cardiac arrhythmia research. *Current Cardiology Reports*, *19*(3), 23.

Goh, K.-I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., & Barabási, A.-L. (2007). The human disease network. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(21), 8685–90.

Goldenberg, I., Mathew, J., Moss, A. J., McNitt, S., Peterson, D. R., Zareba, W., … Morray, B. (2006). Corrected QT variability in serial electrocardiograms in long QT syndrome. The Importance of the Maximum Corrected QT for Risk Stratification. *Journal of the American College of Cardiology*, *48*(5), 1047–1052.

Goldsmith, E. C., Hoffman, A., Morales, M. O., Potts, J. D., Price, R. L., McFadden, A., … Borg, T. K. (2004). Organization of fibroblasts in the heart. *Developmental Dynamics*, *230*(4), 787–794.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*, *17*(6), 333–351.

Graichen, R., Xu, X., Braam, S. R., Balakrishnan, T., Norfiza, S., Sieh, S., … Davidson, B. P. (2008). Enhanced cardiomyogenesis of human embryonic stem cells by a small molecular inhibitor of p38 MAPK. *Differentiation*, *76*(4), 357–370.

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., … Stratton, M. R. (2007). Patterns of somatic mutation in human cancer genomes. *Nature*, *446*(7132), 153–158.

Griggs, R. C., Batshaw, M., Dunkle, M., Gopal-Srivastava, R., Kaye, E., Krischer, J., … Merkel, P. A. (2009). Clinical research for rare disease: opportunities, challenges, and solutions. *Molecular Genetics and Metabolism*, *96*(1), 20–26.

Gromyko, D., Arnesen, T., Ryningen, A., Varhaug, J. E., & Lillehaug, J. R. (2010). Depletion of the human Nα-terminal acetyltransferase A induces p53-dependent apoptosis and p53-independent growth inhibition. *International Journal of Cancer*, *127*(12), 2777–89.

Gymrek, M., Golan, D., Rosset, S., & Erlich, Y. (2012). lobSTR: A short tandem repeat profiler for personal genomes. *Genome Research*, *22*(6), 1154–62.

Hamlin, R. L., & Altschuld, R. A. (2011). Extrapolation from mouse to man. *Circulation. Cardiovascular Imaging*, *4*(1), 2–4.

Hanson, E. H., Imperatore, G., & Burke, W. (2001). HFE Gene and Hereditary Hemochromatosis: a HuGE review. *American Journal of Epidemiology*, *154*(3), 193–206.

Hansson, J., Rafiee, M. R., Reiland, S., Polo, J. M., Gehring, J., Okawa, S., … Krijgsveld, J. (2012). Highly coordinated proteome dynamics during reprogramming of somatic cells to pluripotency. *Cell Reports*, *2*(6), 1579–92.

Hartley, B. J., & Brennand, K. J. (2016). Neural organoids for disease phenotyping, drug screening and developmental biology studies. *Neurochemistry International*, pii: S0197-0186(16)30370-9. [Epub ahead of print].

Hedera, P., & Gorski, J. L. (2003). Oculo-facio-cardio-dental syndrome: skewed X chromosome inactivation in mother and daughter suggest X-linked dominant Inheritance. *American Journal of Medical Genetics. Part A*, *123A*(3), 261–6.

Helbig, A. O., Gauci, S., Raijmakers, R., van Breukelen, B., Slijper, M., Mohammed, S., & Heck, A. J. R. (2010). Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome. *Molecular & Cellular Proteomics*, *9*(5), 928–939.

Hellman-Aharony, S., Smirin-Yosef, P., Halevy, A., Pasmanik-Chor, M., Yeheskel, A., Har-Zahav, A., … Basel-Vanagaite, L. (2013). Microcephaly thin corpus callosum intellectual disability syndrome caused by mutated TAF2. *Pediatr Neurol*, *49*(6), 411–416.

Hendriks, W. T., Warren, C. R., & Cowan, C. A. (2016). Genome editing in human pluripotent stem cells: approaches, pitfalls, and solutions. *Cell Stem Cell*, *18*(1), 53–65.

Herzfeld, T., Nolte, D., Grznarova, M., Hofmann, A., Schultze, J. L., & Muller, U. (2013). X-linked dystonia parkinsonism syndrome (XDP, lubag): disease-specific sequence change DSC3 in TAF1/DYT3 affects genes in vesicular transport and dopamine metabolism. *Hum Mol Genet*, *22*(5), 941–951.

Hetherington, S., McGuirk, S., Powell, G., Cutrell, A., Naderer, O., Spreen, B., … Steel, H.

(2001). Hypersensitivity reactions during therapy with the nucleoside reverse transcriptase inhibitor abacavir. *Clinical Therapeutics*, *23*(10), 1603–14.

Highnam, G., Wang, J. J., Kusler, D., Zook, J., Vijayan, V., Leibovich, N., & Mittelman, D. (2015). An analytical framework for optimizing variant discovery from personal genomes. *Nature Communications*, *6*, 6275.

Hill, B. G., Benavides, G. A., Jr, J. R. L., Ballinger, S., Italia, L. D., Zhang, J., & Darley-usmar, V. M. (2012). Integration of cellular bioenergetics with mitochondrial quality control and autophagy. *Biological Chemistry*, *393*(12), 1485–1512.

Hilton, I. B., & Gersbach, C. A. (2015). Enabling functional genomics with genome engineering. *Genome Research*, *25*(10), 1442–1455.

Hirt, M. N., Hansen, A., & Eschenhagen, T. (2014). Cardiac tissue engineering: state of the art. *Circulation Research*, *114*(2), 354–367.

Hochedlinger, K., & Jaenisch, R. (2015). Induced pluripotency and epigenetic reprogramming. *Cold Spring Harbor Perspectives in Biology*, *7*(12), pii: a019448. [Epub ahead of print].

Hockemeyer, D., & Jaenisch, R. (2016). Induced pluripotent stem cells meet genome editing. *Cell Stem Cell*, *18*(5), 573–586.

Hoischen, A., van Bon, B. W. M., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., … Veltman, J. A. (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nature Genetics*, *42*(6), 483–485.

Holmes, W. M., Mannakee, B. K., Gutenkunst, R. N., & Serio, T. R. (2014). Loss of amino-terminal acetylation suppresses a prion phenotype by modulating global protein folding. *Nature Communications*, *5*, 4383.

Horton, J. L., Martin, O. J., Lai, L., Riley, N. M., Richards, A. L., Vega, R. B., … Kelly, D. P. (2016). Mitochondrial protein hyperacetylation in the failing heart. *JCI Insight*, *1*(2), 1–14.

Hotta, A., & Yamanaka, S. (2015). From genomics to gene therapy: induced pluripotent stem cells meet genome editing. *Annual Review of Genetics*, (September), 1–24.

Hu, H., Haas, S. A., Chelly, J., Van Esch, H., Raynaud, M., de Brouwer, A. P., … Kalscheuer, V. M. (2015). X-exome sequencing of 405 unresolved families identifies seven novel intellectual disability genes. *Molecular Psychiatry*.

Hua, K.-T., Tan, C.-T., Johansson, G., Lee, J.-M., Yang, P.-W., Lu, H.-Y., … Kuo, M.-L. (2011). N-α-acetyltransferase 10 protein suppresses cancer cell metastasis by binding PIX proteins and inhibiting Cdc42/Rac1 activity. *Cancer Cell*, *19*(2), 218–31.

Huang, B. E., Mulyasasmita, W., & Rajagopal, G. (2016). The path from big data to precision medicine. *Expert Review of Precision Medicine and Drug Development*, *1*(2), 129–143.

Hughes, A. R., Brothers, C. H., Mosteller, M., Spreen, W. R., & Burns, D. K. (2009). Genetic association studies to detect adverse drug reactions: abacavir hypersensitivity as an example.

Pharmacogenomics, *10*(2), 225–233.

Hung, S. S. C., Khan, S., Lo, C. Y., Hewitt, A. W., & Wong, R. C. B. (2017). Drug discovery using induced pluripotent stem cell models of neurodegenerative and ocular diseases. *Pharmacology & Therapeutics*, pii: S0163-7258(17)30040-2. [Epub ahead of print].

Hunt, K. A., Zhernakova, A., Turner, G., Heap, G. A. R., Franke, L., Bruinenberg, M., … van Heel, D. A. (2008). Newly identified genetic risk variants for celiac disease related to the immune response. *Nature Genetics*, *40*(4), 395–402.

Hwang, C.-S., Shemorry, A., & Varshavsky, A. (2010). N-terminal acetylation of cellular proteins creates specific degradation signals. *Science*, *327*(5968), 973–7.

Hwang, S., Kim, E., Lee, I., & Marcotte, E. M. (2015). Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Scientific Reports*, *5*(December), 17875.

Im, H., Lee, H., & Castro, C. M. (2016). Challenges influencing next generation technologies for precision medicine. *Expert Review of Precision Medicine and Drug Development*, *1*(2), 121–123.

Indo, Y. (2001). Molecular basis of congenital insensitivity to pain with anhidrosis (CIPA): mutations and polymorphisms in TRKA (NTRK1) gene encoding the receptor tyrosine kinase for nerve growth factor. *Human Mutation*, *18*(6), 462–471.

Indo, Y., Tsuruta, M., Hayashida, Y., Karim, M. A., Ohta, K., Kawano, T., … Matsuda, I. (1996). Mutations in the TRKA/NGF receptor gene in patients with congenital insensitivity to pain with anhidrosis. *Nat Genet*, *13*(4), 485–488.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. *Nature*, *431*(7011), 931–45.

International Multiple Sclerosis Genetics Consortium, S., Wellcome Trust Case Control Consortium 2, G., Sawcer, S., Hellenthal, G., Pirinen, M., Spencer, C. C. A., … Compston, A. (2011). Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature*, *476*(7359), 214–9.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., … Wigler, M. (2012). *De novo* gene disruptions in children on the autistic spectrum. *Neuron*, *74*(2), 285–299.

Ipata, P. L., & Balestri, F. (2012). Glycogen as a fuel: metabolic interaction between glycogen and ATP catabolism in oxygen-independent muscle contraction. *Metabolomics*, *8*(4), 736–741.

Itzhaki, I., Maizels, L., Huber, I., Zwi-dantsis, L., Caspi, O., Winterstern, A., … Gepstein, L. (2011). Modelling the long QT syndrome with induced pluripotent stem cells. *Nature*, *471*(7337), 225–229.

Jambaldorj, J., Makino, S., Munkhbat, B., & Tamiya, G. (2012). Sustained expression of a neuron-specific isoform of the Taf1 gene in development stages and aging in mice. *Biochem*

Biophys Res Commun, *425*(2), 273–277.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., & Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife*, *2*, e00471.

John, R. M., Tedrow, U. B., Koplan, B. A., Albert, C. M., Epstein, L. M., Sweeney, M. O., … Stevenson, W. G. (2012). Ventricular arrhythmias and sudden cardiac death. *Lancet*, *380*(9852), 1520–9.

Johnson, J. O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V. M., Trojanowski, J. Q., … Traynor, B. J. (2010). Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron*, *68*(5), 857–64.

Jung, C. B., Moretti, A., Mederos y Schnitzler, M., Iop, L., Storch, U., Bellin, M., … Laugwitz, K.-L. (2012). Dantrolene rescues arrhythmogenic RYR2 defect in a patient-specific stem cell model of catecholaminergic polymorphic ventricular tachycardia. *EMBO Molecular Medicine*, *4*(3), 180–91.

Kalvik, T. V, & Arnesen, T. (2013). Protein N-terminal acetyltransferases in cancer. *Oncogene*, *32*(3), 269–76.

Kan, Z., Zheng, H., Liu, X., Li, S., Barber, T. D., Gong, Z., … Mao, M. (2013). Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Research*, *23*(9), 1422–33.

Karagiannis, P., Eto, K., Bae, S., Park, J., Kim, J.-S., Chong, J. J. H., … Tsumaki, N. (2016). Ten years of induced pluripotency: from basic mechanisms to therapeutic applications. *Development (Cambridge, England)*, *143*(12), 2039–43.

Kidd, J. M., Sampas, N., Antonacci, F., Graves, T., Fulton, R., Hayden, H. S., … Eichler, E. E. (2010). Characterization of missing human genome sequences and copy-number polymorphic insertions. *Nature Methods*, *7*(5), 365–71.

Kim, C., Wong, J., Wen, J., Wang, S., Wang, C., Spiering, S., … Chen, H.-S. V. (2013). Studying arrhythmogenic right ventricular dysplasia with patient-specific iPSCs. *Nature*, *494*(7435), 105–110.

Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., … Pandey, A. (2014). A draft map of the human proteome. *Nature*, *509*(7502), 575–581.

Kléber, A. G., & Rudy, Y. (2004). Basic mechanisms of cardiac impulse propagation and associated arrhythmias. *Physiological Reviews*, *84*(2), 431–488.

Klimas, A., Ambrosi, C. M., Yu, J., Williams, J. C., Bien, H., & Entcheva, E. (2016). OptoDyCE as an automated system for high-throughput all-optical dynamic cardiac electrophysiology. *Nature Communications*, *7*, 11542.

Knudsen, G. P. S., Neilson, T. C. S., Pedersen, J., Kerr, A., Schwartz, M., Hulten, M., … Orstavik, K. H. (2006). Increased skewing of X chromosome inactivation in Rett syndrome patients and their mothers. *European Journal of Human Genetics*, *14*(11), 1189–94.

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). The Next-generation sequencing revolution and its impact on genomics. *Cell*, *155*(1), 27–38.

Kohane, B. I. S. (2015). Ten things we have to do to achieve precision medicine. *Science*, *349*(6243), 37–8.

Köhler, S., Bauer, S., Horn, D., & Robinson, P. N. (2008). Walking the interactome for prioritization of candidate disease genes. *The American Journal of Human Genetics*, *82*(4), 949–958.

Kohler, S., Vasilevsky, N. A., Engelstad, M., Foster, E., McMurry, J., Ayme, S., … Robinson, P. N. (2017). The Human Phenotype Ontology in 2017. *Nucleic Acids Res*, *45*(D1), D865–D876.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, *533*(7603), 420–4.

Korngiebel, D. M., Thummel, K. E., & Burke, W. (2017). Implementing precision medicine: the ethical challenges. *Trends in Pharmacological Sciences*, *38*(1), 8–14.

Krahn, A. D., Healey, J. S., Chauhan, V., Birnie, D. H., Simpson, C. S., Champagne, J., … Gollob, M. H. (2009). Systematic assessment of patients with unexplained cardiac arrest: cardiac arrest survivors with preserved ejection fraction registry (CASPER). *Circulation*, *120*(4), 278–285.

Kuo, H.-P., Lee, D.-F., Chen, C.-T., Liu, M., Chou, C.-K., Lee, H.-J., … Hung, M.-C. (2010). ARD1 stabilization of TSC2 suppresses tumorigenesis through the mTOR signaling pathway. *Science Signaling*, *3*(108), ra9.

La Du, B. N., Zannoni, V. G., Laster, L., & Seegmiller, J. E. (1958). The nature of the defect in tyrosine metabolism in alcaptonuria. *The Journal of Biological Chemistry*, *230*(1), 251–60.

Laflamme, M. A., Chen, K. Y., Naumova, A. V, Muskheli, V., Fugate, J. A., Dupras, S. K., … Murry, C. E. (2007). Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts. *Nature Biotechnology*, *25*(9), 1015–24.

Lam, H. Y., Clark, M. J., Chen, R., Chen, R., Natsoulis, G., O'Huallachain, M., … Snyder, M. (2012). Performance comparison of whole-genome sequencing platforms. *Nature Biotechnology*, *30*(1), 78–82.

Landrum, M. J., Lee, J. M., Riley, G. R., Jang, W., Rubinstein, W. S., Church, D. M., & Maglott, D. R. (2014). ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, *42*(Database issue), D980-5.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–9.

Lee, C.-Y., Chiu, Y.-C., Wang, L.-B., Kuo, Y.-L., Chuang, E. Y., Lai, L.-C., & Tsai, M.-H. (2013). Common applications of next-generation sequencing technologies in genomic research. *Translational Cancer Research*, *2*(1), 33–45.

Lee, C., Ou, D. S., Lee, S., Chang, L., Lin, R., Li, Y., … Juan, L. (2010). hNaa10p contributes to tumorigenesis by facilitating DNMT1-mediated tumor suppressor gene silencing. *The Journal of Clinical Investigation*, *120*(8), 2920–30.

Lee, W.-P., Stromberg, M. P., Ward, A., Stewart, C., Garrison, E. P., & Marth, G. T. (2014). MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. *PloS One*, *9*(3), e90581.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv E-Prints*.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Subgroup, G. P. D. P. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, H. O., Zhu, Y. F., Asakawa, M., Kuma, H., Hirata, T., Ueda, Y., … Hasegawa, M. (2000). A cytoplasmic RNA vector derived from nontransmissible Sendai virus with efficient gene transfer and expression. *Journal of Virology*, *74*(14), 6564–9.

Li, M., & Belmonte, J. C. I. (2017). Ground rules of the pluripotency gene regulatory network. *Nature Reviews Genetics*, *18*(3), 180–191.

Li, R., Li, Y., Fang, X., Yang, H., Wang, J. J., Kristiansen, K., & Wang, J. J. (2009). SNP detection for massively parallel whole-Genome Researchequencing. *Genome Research*, *19*(6), 1124–1132.

Li, R., Li, Y., Zheng, H., Luo, R., Zhu, H., Li, Q., … Wang, J. (2010). Building the sequence map of the human pan-genome. *Nature Biotechnology*, *28*(1), 57–63.

Lian, X., Hsiao, C., Wilson, G., Zhu, K., Hazeltine, L. B., Azarin, S. M., & Raval, K. K. (2012). Robust cardiomyocyte differentiation from human pluripotent stem cells via temporal modulation of canonical Wnt signaling. *Proceedings of the National Academy of Sciences*, *109*(27), E1848-57.

Lian, X., Zhang, J., Azarin, S. M., Zhu, K., Hazeltine, L. B., Bao, X., … Palecek, S. P. (2012). Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. *Nat Protoc*, *8*(1), 162–75.

Liao, P., & Soong, T. W. (2010). $Ca_V1.2$ channelopathies: from arrhythmias to autism, bipolar disorder, and immunodeficiency. *Pflügers Archiv : European Journal of Physiology*, *460*(2), 353–9.

Lim, J.-H., Chun, Y.-S., & Park, J.-W. (2008). Hypoxia-inducible factor-1alpha obstructs a Wnt

signaling pathway by inhibiting the hARD1-mediated activation of beta-catenin. *Cancer Research*, *68*(13), 5177–84.

Lim, J.-H., Park, J.-W., & Chun, Y.-S. (2006). Human arrest defective 1 acetylates and activates beta-catenin, promoting lung cancer cell proliferation. *Cancer Research*, *66*(22), 10677–82.

Limpitikul, W. B., Dick, I. E., Joshi-Mukherjee, R., Overgaard, M. T., George, A. L., & Yue, D. T. (2014). Calmodulin mutations associated with long QT syndrome prevent inactivation of cardiac L-type Ca^{2+} currents and promote proarrhythmic behavior in ventricular myocytes. *Journal of Molecular and Cellular Cardiology*, *74*, 115–124.

Lindeman, N. I., Cagle, P. T., Beasley, M. B., Chitale, D. A., Dacic, S., Giaccone, G., … Ladanyi, M. (2013). Molecular testing guideline for selection of lung cancer patients for egfr and alk tyrosine kinase inhibitors: guideline from the college of american pathologists, international association for the study of lung cancer, and association for molecular pathology. *Journal of Thoracic Oncology*, *8*(7), 823–859.

Linē, A., Stengrēvics, A., Slucka, Z., Li, G., Jankevics, E., & Rees, R. C. (2002). Serological identification and expression analysis of gastric cancer-associated genes. *British Journal of Cancer*, *86*(11), 1824–1830.

Liszczak, G., Goldberg, J. M., Foyn, H., Petersson, E. J., Arnesen, T., & Marmorstein, R. (2013). Molecular basis for N-terminal acetylation by the heterodimeric NatA complex. *Nature Structural & Molecular Biology*, *20*(9), 1098–105.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., … Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, *2012*, *article ID 251364*.

Liu, X., Han, S., Wang, Z., Gelernter, J., & Yang, B.-Z. (2013). Variant callers for next-generation sequencing data: a comparison study. *PLoS ONE*, *8*(9), e75619.

Lopaschuk, G. D. (2016). Metabolic modulators in heart disease – past, present and future. *Canadian Journal of Cardiology*, pii: S0828-282X(16)31172-2. [Epub ahead of print].

Lopes-Pacheco, M. (2016). CFTR modulators: shedding light on precision medicine for cystic fibrosis. *Frontiers in Pharmacology*, *7*, 275.

Lowe, W. L., & Reddy, T. E. (2015). Genomic approaches for understanding the genetics of complex disease. *Genome Research*, *25*(10), 1432–41.

Lundy, S. D., Zhu, W.-Z., Regnier, M., & Laflamme, M. A. (2013). Structural and functional maturation of cardiomyocytes derived from human pluripotent stem cells. *Stem Cells and Development*, *22*(14), 1991–2002.

Luthra, R., Chen, H., Roy-Chowdhuri, S., & Singh, R. R. (2015). Next-generation sequencing in clinical molecular diagnostics of cancer: advantages and challenges. *Cancers*, *7*(4), 2023–2036.

Lyon, G. J. (2012). Personalized medicine: Bring clinical standards to human-genetics research.

Nature, *482*(7385), 300–301.

Lyon, G. J., & Segal, J. P. (2013). Practical, ethical and regulatory considerations for the evolving medical and research genomics landscape. *Applied and Translational Genomics*, *2*(1), 34–40.

LYON, M. F. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature*, *190*, 372–3.

Ma, D., Wei, H., Lu, J., Ho, S., Zhang, G., Sun, X., … Liew, R. (2013). Generation of patient-specific induced pluripotent stem cell-derived cardiomyocytes as a cellular model of arrhythmogenic right ventricular cardiomyopathy. *European Heart Journal*, *34*(15), 1122–33.

Mackay, T. F. C., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet*, *10*(8), 565–577.

Makino, S., Kaji, R., Ando, S., Tomizawa, M., Yasuno, K., Goto, S., … Tamiya, G. (2007). Reduced neuron-specific expression of the TAF1 gene is associated with X-linked dystonia-parkinsonism. *American Journal of Human Genetics*, *80*(3), 393–406.

Malan, D., Friedrichs, S., Fleischmann, B. K., & Sasse, P. (2011). Cardiomyocytes obtained from induced pluripotent stem cells with long-QT syndrome 3 recapitulate typical disease-specific features in vitro. *Circulation Research*, *109*(8), 841–7.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., … Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), 823–826.

Mallal, S., Phillips, E., Carosi, G., Molina, J.-M., Workman, C., Tomažič, J., … PREDICT-1 Study Team. (2008). HLA-B*5701 screening for hypersensitivity to abacavir. *New England Journal of Medicine*, *358*(6), 568–579.

Manning, A. L., & Dyson, N. J. (2011). pRB, a tumor suppressor with a stabilizing presence. *Trends in Cell Biology*, *21*(8), 433–441.

Mardis, E. R. (2013). Next-generation sequencing platforms. *Annual Review of Analytical Chemistry (Palo Alto, Calif.)*, *6*, 287–303.

Mardis, E. R. (2017). DNA sequencing technologies: 2006–2016. *Nature Protocols*, *12*(2), 213–218.

Marionneau, C., & Abriel, H. (2015). Regulation of the cardiac Na^+ channel $Na_v1.5$ by post-translational modifications. *Journal of Molecular and Cellular Cardiology*, *82*, 36–47.

Marshall, N. J., Goodwin, C. J., & Holt, S. J. (1995). A critical assessment of the use of microculture tetrazolium assays to measure cell growth and function. *Growth Regulation*, *5*(2), 69–84.

Martin, D. T., Gendron, R. L., Jarzembowski, J. A., Perry, A., Collins, M. H., Pushpanathan, C., … Paradis, H. (2007). Tubedown expression correlates with the differentiation status

and aggressiveness of neuroblastic tumors. *Clinical Cancer Research*, *13*(5), 1480–7.

Martin, M. A., & Kroetz, D. L. (2013). Abacavir pharmacogenetics - from initial reports to standard of care. *Pharmacotherapy*, *33*(7), 765–775.

Martiniano, S. L., Sagel, S. D., & Zemanick, E. T. (2016). Cystic fibrosis: a model system for precision medicine. *Current Opinion in Pediatrics*, *28*(3), 312–317.

Massouras, A., Waszak, S. M., Albarca-Aguilera, M., Hens, K., Holcombe, W., Ayroles, J. F., … Deplancke, B. (2012). Genomic variation and its Impact on gene expression in *Drosophila melanogaster*. *PLoS Genet*, *8*(11), e1003055.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303.

McLaren, C. E., Emond, M. J., Subramaniam, V. N., Phatak, P. D., Barton, J. C., Adams, P. C., … McLaren, G. D. (2015). Exome sequencing in HFE C282Y homozygous men with extreme phenotypes identifies a GNPAT variant associated with severe iron overload. *Hepatology*.

Medeiros-Domingo, A., Bhuiyan, Z. A., Tester, D. J., Hofman, N., Bikker, H., van Tintelen, J. P., … Ackerman, M. J. (2009). The RYR2-encoded ryanodine receptor/calcium release channel in patients diagnosed previously with either catecholaminergic polymorphic ventricular tachycardia or genotype negative, exercise-induced long QT syndrome: a comprehensive open reading frame mutational analysis. *Journal of the American College of Cardiology*, *54*(22), 2065–2074.

Meienberg, J., Zerjavic, K., Keller, I., Okoniewski, M., Patrignani, A., Ludin, K., … Matyas, G. (2015). New insights into the performance of human whole-exome capture platforms. *Nucleic Acids Research*, *43*(11), e76.

Metzker, M. L. (2010). Sequencing technologies - the next generation. *Nature Reviews. Genetics*, *11*(1), 31–46.

Mitchell, K. J. (2012). What is complex about complex disorders? *Genome Biology*, *13*(1), 237.

Moczulski, D. K., Grzeszczak, W., & Gawlik, B. (2001). Role of hemochromatosis C282Y and H63D mutations in HFE gene in development of type 2 diabetes and diabetic nephropathy. *Diabetes Care*, *24*(7), 1187–1191.

Monda, J. K., Scott, D. C., Miller, D. J., Lydeard, J., King, D., Harper, J. W., … Schulman, B. A. (2013). Structural conservation of distinctive N-terminal acetylation-dependent interactions across a family of mammalian NEDD8 ligation enzymes. *Structure (London, England : 1993)*, *21*(1), 42–53.

Moretti, A., Bellin, M., Welling, A., Jung, C. B., Lam, J. T., Bott-Flügel, L., … Laugwitz, K.-L. (2010). Patient-specific induced pluripotent stem-cell models for long-QT syndrome. *The New England Journal of Medicine*, *363*(15), 1397–409.

Moss, A. J., Schwartz, P. J., Crampton, R. S., Tzivoni, D., Locati, E. H., MacCluer, J., … Garson, A. (1991). The long QT syndrome. Prospective longitudinal study of 328 families. *Circulation*, *84*(3), 1136–1144.

Mozos, I., & Caraba, A. (2015). Electrocardiographic predictors of cardiovascular mortality. *Disease Markers*, *2015*.

Mullen, J. R., Kaynel, P. S., Moerschell, R. P., Tsunasawa, S., Gribskovl, M., Colavito-shepanskil, M., … Sternglanz, R. (1989). Identification and characterization of genes and mutants for an N-terminal acetyltransferase from yeast. *The EMBO Journal*, *8*(7), 2067–2075.

Mummery, C. L., Zhang, J., Ng, E. S., Elliott, D. A., Elefanty, A. G., & Kamp, T. J. (2012). Differentiation of human embryonic stem cells and induced pluripotent stem cells to cardiomyocytes: a methods overview. *Circulation Research*, *111*(3), 344–358.

Musunuru, K. (2013). Genome editing of human pluripotent stem cells to generate human cellular disease models. *Disease Models & Mechanisms*, *6*(4), 896–904.

Muthuswamy, S. K. (2017). Bringing together the organoid field: from early beginnings to the road ahead. *Development*, *144*(6), 963–967.

Myklebust, L. M., Van Damme, P., Støve, S. I., Dörfel, M. J., Abboud, A., Kalvik, T. V., … Arnesen, T. (2014). Biochemical and cellular analysis of Ogden syndrome reveals downstream Nt-acetylation defects. *Human Molecular Genetics*, *24*(7), 1956–1976.

Naito, A. T., Shiojima, I., Akazawa, H., Hidaka, K., Morisaki, T., Kikuchi, A., & Komuro, I. (2006). Developmental stage-specific biphasic roles of Wnt/beta-catenin signaling in cardiomyogenesis and hematopoiesis. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(52), 19812–7.

Narzisi, G., O'Rawe, J. A., Iossifov, I., Fang, H., Lee, Y.-H., Wang, Z., … Schatz, M. C. (2014). Accurate *de novo* and transmitted indel detection in exome-capture data using microassembly. *Nature Methods*, *11*(10), 1033–6.

Nevins, J. R. (2001). The Rb/E2F pathway and cancer. *Human Molecular Genetics*, *10*(7), 699–703.

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., … Shendure, J. (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, *461*(7261), 272–276.

Nieves Calatrava, D., Calle-Martín, Ó. de la, Iribarren-Loyarte, J. A., Rivero-Román, A., García-Bujalance, L., Pérez-Escolano, I., & Brosa-Riestra, M. (2010). Cost-effectiveness analysis of HLA-B*5701 typing in the prevention of hypersensitivity to abacavir in HIV+ patients in Spain. *Enfermedades Infecciosas Y Microbiología Clínica*, *28*(9), 590–595.

Nolasco, J. B., & Dahlen, R. W. (1968). A graphic method for the study of alternation in cardiac action potentials. *The Journal of Applied Physiology*, *25*(2), 191–6.

Norcliffe-Kaufmann, L., Slaugenhaupt, S. A., & Kaufmann, H. (2016). Familial dysautonomia: history, genotype, phenotype and translational research. *Progress in Neurobiology*, pii: S0301-0082(16)30020-X.

Novak, A., Barad, L., Zeevi-Levin, N., Shick, R., Shtrichman, R., Lorber, A., … Binah, O. (2012). Cardiomyocytes generated from CPVTD307H patients are arrhythmogenic in response to β-adrenergic stimulation. *Journal of Cellular and Molecular Medicine*, *16*(3), 468–82.

O'Halloran, D. M. (2015). PrimerView: high-throughput primer design and visualization. *Source Code for Biology and Medicine*, *10*, 8.

O'Halloran, D. M., Leitner, T., Blocker, H., Marky, L. A., & Lipman, D. J. (2016). PrimerMapper: high throughput primer design and graphical assembly for PCR and SNP detection. *Scientific Reports*, *6*(1), 20631.

O'Rawe, J. A., Wu, Y., Dorfel, M. J., Rope, A. F., Au, P. Y. B., Parboosingh, J. S., … Lyon, G. J. (2015). TAF1 variants are associated with dysmorphic features, intellectual disability, and neurological manifestations. *American Journal of Human Genetics*, *97*(6), 922–932.

O'Rawe, J., Jiang, T., Sun, G., Wu, Y., Wang, W., Hu, J., … Lyon, G. J. (2013). Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Medicine*, *5*(3), 28.

O'Rawe, J., Wu, Y., Rope, A., Jimenez Barrón, L. T., Swensen, J., Fang, H., … Lyon, G. (2015). A variant in TAF1 is associated with a new syndrome with severe intellectual disability and characteristic dysmorphic features. *bioRxiv*, https://doi.org/10.1101/014050.

Office of the Press Secretary. (2015). FACT SHEET: President Obama's Precision Medicine Initiative | whitehouse.gov. Retrieved January 30, 2015, from https://obamawhitehouse.archives.gov/the-press-office/2015/01/30/fact-sheet-president-obama-s-precision-medicine-initiative.

Office of the Press Secretary. (2016). FACT SHEET: Investing in the National Cancer Moonshot | whitehouse.gov. Retrieved February 01, 2016, from https://obamawhitehouse.archives.gov/the-press-office/2016/02/01/fact-sheet-investing-national-cancer-moonshot.

Ohnuki, M., & Takahashi, K. (2015). Present and future challenges of induced pluripotent stem cells. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1680), 20140367.

Okata, S., Yuasa, S., Suzuki, T., Ito, S., Makita, N., Yoshida, T., … Fukuda, K. (2016). Embryonic type Na^+ channel β-subunit, SCN3B masks the disease phenotype of Brugada syndrome. *Scientific Reports*, *6*(1), 34198.

Ombrello, M. J., Remmers, E. F., Sun, G., Freeman, A. F., Datta, S., Torabi-Parizi, P., … Milner, J. D. (2012). Cold urticaria, immunodeficiency, and autoimmunity related to PLCG2 deletions. *The New England Journal of Medicine*, *366*(4), 330–8.

Osler, W. (1892). *The principles and practice of medicine: designed for the use of practitioners and students of medicine* (First). Edinburgh & London : Young J. Pentland.

Oti, M., & Brunner, H. (2006). The modular nature of genetic diseases. *Clinical Genetics*, *71*(1), 1–11.

Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., … Tessier-Lavigne, M. (2016). Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*, *533*(7601), 125–9.

Pelucchi, S., Mariani, R., Calza, S., Fracanzani, A. L., Modignani, G. L., Bertola, F., … Piperno, A. (2012). CYBRD1 as a modifier gene that modulates iron phenotype in HFE p.C282Y homozygous patients. *Haematologica*, *97*(12), 1818–1825.

Peng, S., Lacerda, A. E., Kirsch, G. E., Brown, A. M., & Bruening-Wright, A. (2010). The action potential and comparative pharmacology of stem cell-derived human cardiomyocytes. *Journal of Pharmacological and Toxicological Methods*, *61*(3), 277–286.

Pennington, K. L., Marr, S. K., Chirn, G. W., & Marr 2nd, M. T. (2013). Holo-TFIID controls the magnitude of a transcription burst and fine-tuning of transcription. *Proc Natl Acad Sci U S A*, *110*(19), 7678–7683.

Perlman, R. L., & Govindaraju, D. R. (2016). Archibald E. Garrod: the father of precision medicine. *Genetics in Medicine*, *18*(11), 1088–1089.

Pettitt, D., Smith, J., Meadows, N., Arshad, Z., Schuh, A., DiGiusto, D., … Brindley, D. (2016). Regulatory barriers to the advancement of precision medicine. *Expert Review of Precision Medicine and Drug Development*, *1*(3), 319–329.

Pietrangelo, A. (2004). Hereditary hemochromatosis — a new look at an old disease. *New England Journal of Medicine*, *350*(23), 2383–2397.

Polevoda, B., Brown, S., Cardillo, T. S., Rigby, S., & Sherman, F. (2008). Yeast Nα-terminal acetyltransferases are associated with ribosomes. *Journal of Cellular Biochemistry*, *103*(2), 492–508.

Popp, B., Støve, S. I., Endele, S., Myklebust, L. M., Hoyer, J., Sticht, H., … Reis, A. (2015). De novo missense mutations in the NAA10 gene cause severe non-syndromic developmental delay in males and females. *The European Journal of Human Genetics*, *23*(5), 602–609.

Prada, C. E., Gonzaga-Jauregui, C., Tannenbaum, R., Penney, S., Lupski, J. R., Hopkin, R. J., & Sutton, V. R. (2014). Clinical utility of whole-exome sequencing in rare diseases: Galactosialidosis. *European Journal of Medical Genetics*, *57*(7), 339–344.

Prakash, V., Moore, M., & Yáñez-Muñoz, R. J. (2016). Current progress in therapeutic gene editing for monogenic diseases. *Molecular Therapy*, *24*(3), 465–474.

Qiu, X., Xiao, X., Li, N., & Li, Y. (2017). Histone deacetylases inhibitors (HDACis) as novel therapeutic application in various clinical diseases. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *72*, 60–72.

Quail, M., Smith, M. E., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., … Salzberg, S. (2012). A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. *BMC Genomics*, *13*(1), 341.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.

Ragni, M. V., Moore, C. G., Bias, V., Key, N. S., Kouides, P. A., & Francis, C. W. (2012). Challenges of rare disease research: Limited patients and competing priorities. *Haemophilia*, *18*(3), 2011–2013.

Raychaudhuri, S., Sinha, M., Mukhopadhyay, D., & Bhattacharyya, N. P. (2008). HYPK, a Huntingtin interacting protein, reduces aggregates and apoptosis induced by N-terminal Huntingtin with 40 glutamines in Neuro2a cells and exhibits chaperone-like activity. *Human Molecular Genetics*, *17*(2), 240–55.

Rehm, H. L., Berg, J. S., Brooks, L. D., Bustamante, C. D., Evans, J. P., Landrum, M. J., … Watson, M. S. (2015). ClinGen — The Clinical Genome Resource. *New England Journal of Medicine*, *372*(23), 2235–2242.

Reinhardt, P., Schmid, B., Burbulla, L. F., Schöndorf, D. C., Wagner, L., Glatza, M., … Sterneckert, J. (2013). Genetic correction of a LRRK2 mutation in human iPSCs links parkinsonian neurodegeneration to ERK-dependent changes in gene expression. *Cell Stem Cell*, *12*(3), 354–67.

Ren, T., Jiang, B., Jin, G., Li, J., Dong, B., Zhang, J., … Shou, C. (2008). Generation of novel monoclonal antibodies and their application for detecting ARD1 expression in colorectal cancer. *Cancer Letters*, *264*(1), 83–92.

Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., … Traynor, B. J. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*, *72*(2), 257–268.

Reppel, M., Boettinger, C., & Hescheler, J. (2004). Beta-adrenergic and muscarinic modulation of human embryonic stem cell-derived cardiomyocytes. *Cellular Physiology and Biochemistry*, *14*(4–6), 187–96.

Rhrissorrakrai, K., Koyama, T., & Parida, L. (2016). Watson for genomics: moving personalized medicine forward. *Trends in Cancer*, *2*(8), 392–395.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., … ACMG Laboratory Quality Assurance Committee. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, *17*(5), 405–24.

Rieber, N., Zapatka, M., Lasitschka, B., Jones, D., Northcott, P., Hutter, B., … Eils, R. (2013). Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies. *PLoS ONE*, *8*(6), e66621.

Rijnbeek, P. R., Witsenburg, M., Schrama, E., Hess, J., & Kors, J. A. (2001). New normal limits for the paediatric electrocardiogram. *European Heart Journal*, *22*(8), 702–711.

Robinson, P. N., Köhler, S., Bauer, S., Seelow, D., Horn, D., & Mundlos, S. (2008). The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *The American Journal of Human Genetics*, *83*(5), 610–615.

Robinson, P. N., & Mundlos, S. (2010). The Human Phenotype Ontology. *Clinical Genetics*, *77*(6), 525–534.

Rooms, L., Reyniers, E., Scheers, S., van Luijk, R., Wauters, J., Van Aerschot, L., … Kooy, R. F. (2006). TBP as a candidate gene for mental retardation in patients with subtelomeric 6q deletions. *Eur J Hum Genet*, *14*(10), 1090–1096.

Rope, A. F., Wang, K., Evjenth, R., Xing, J., Johnston, J. J., Swensen, J. J., … Lyon, G. J. (2011). Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency. *American Journal of Human Genetics*, *89*(1), 28–43.

Rosenfeld, J. A., Mason, C. E., & Smith, T. M. (2012). Limitations of the human reference genome for personalized genomics. *PLoS One*, *7*(7), e40294.

Roses, A. D. (2000). Pharmacogenetics and the practice of medicine. *Nature*, *405*(6788), 857–865.

Roy, S., LaFramboise, W. A., Nikiforov, Y. E., Nikiforova, M. N., Routbort, M. J., Pfeifer, J., … Pantanowitz, L. (2016). Next-generation sequencing informatics: challenges and strategies for implementation in a clinical environment. *Archives of Pathology & Laboratory Medicine*, *140*(9), 958–975.

Sabir, I. N., Li, L. M., Grace, A. A., & Huang, C. L. H. (2008). Restitution analysis of alternans and its relationship to arrhythmogenicity in hypokalaemic Langendorff-perfused murine hearts. *Pflugers Archiv European Journal of Physiology*, *455*(4), 653–666.

Sako, W., Morigaki, R., Kaji, R., Tooyama, I., Okita, S., Kitazato, K., … Goto, S. (2011). Identification and localization of a neuron-specific isoform of TAF1 in rat brain: implications for neuropathology of DYT3 dystonia. *Neuroscience*, *189*, 100–107.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–7.

Saunders, C. J., Miller, N. A., Soden, S. E., Dinwiddie, D. L., Noll, A., Alnadi, N. A., … Kingsmore, S. F. (2012). Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Science Translational Medicine*, *4*(154), 154ra135-154ra135.

Saunier, C., Støve, S. I., Popp, B., Gérard, B., Blenski, M., AhMew, N., … Zweier, C. (2016). Expanding the phenotype associated with NAA10-related N-terminal acetylation deficiency. *Human Mutation*, *37*(8), 755–64.

Sayed, N., Liu, C., & Wu, J. C. (2016). Translation of human-induced pluripotent stem cells. *Journal of the American College of Cardiology*, *67*(18), 2161–2176.

Scaglione, K. M., Basrur, V., Ashraf, N. S., Konen, J. R., Elenitoba-Johnson, K. S. J., Todi, S. V, & Paulson, H. L. (2013). The ubiquitin-conjugating enzyme (E2) Ube2w ubiquitinates the N terminus of substrates. *The Journal of Biological Chemistry*, *288*(26), 18784–8.

Schiza, V., Molina-Serrano, D., Kyriakou, D., Hadjiantoniou, A., & Kirmizis, A. (2013). N-alpha-terminal acetylation of histone H4 regulates arginine methylation and ribosomal DNA silencing. *PLoS Genetics*, *9*(9), e1003805.

Schwartz, P. J., & Crotti, L. (2011). QTc behavior during exercise and genetic testing for the long-qt syndrome. *Circulation*, *124*(20), 2181–2184.

Schwartz, P. J., Priori, S. G., Locati, E. H., Napolitano, C., Cantù, F., Towbin, J. A., … Colatsky, T. J. (1995). Long QT syndrome patients with mutations of the SCN5A and HERG genes have differential responses to Na^+ channel blockade and to increases in heart rate. Implications for gene-specific therapy. *Circulation*, *92*(12), 3381–6.

Schwartz, P. J., Stramba-Badiale, M., Crotti, L., Pedrazzini, M., Besana, A., Bosi, G., … Spazzolini, C. (2009). Prevalence of the congenital long-qt syndrome. *Circulation*, *120*(18), 1761–1767.

Shatzky, S., Moses, S., Levy, J., Pinsk, V., Hershkovitz, E., Herzog, L., … Parvari, R. (2000). Congenital insensitivity to pain with anhidrosis (CIPA) in Israeli-Bedouins: genetic heterogeneity, novel mutations in the TRKA/NGF receptor gene, clinical findings, and results of nerve conduction studies. *American Journal of Medical Genetics*, *92*(5), 353–360.

Shemorry, A., Hwang, C.-S., & Varshavsky, A. (2013). Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. *Molecular Cell*, *50*(4), 540–51.

Shi, Y., Inoue, H., Wu, J. C., & Yamanaka, S. (2016). Induced pluripotent stem cell technology: a decade of progress. *Nature Reviews Drug Discovery*, *16*(2), 115–130.

Shin, D. H., Chun, Y.-S., Lee, K.-H., Shin, H.-W., & Park, J.-W. (2009). Arrest defective-1 controls tumor cell behavior by acetylating myosin light chain kinase. *PloS One*, *4*(10), e7451.

Silva, R. D., & Martinho, R. G. (2015). Developmental roles of protein N-terminal acetylation. *Proteomics*, *15*(14), 2402–9.

Siravegna, G., Marsoni, S., Siena, S., & Bardelli, A. (2017). Integrating liquid biopsies into the management of cancer. *Nature Reviews Clinical Oncology*, doi: 10.1038/nrclinonc.2017.14. [Epub ahead of pri.

Solomon, B. D. (2014). Obstacles and opportunities for the future of genomic medicine. *Molecular Genetics & Genomic Medicine*, *2*(3), 205–9.

Solomon, S. (2015). *Ethical Challenges to Next-Generation Sequencing*. *Clinical Genomics*.

Elsevier Inc.

Søndergaard, M. T., Tian, X., Liu, Y., Wang, R., Chazin, W. J., Chen, S. R. W., & Overgaard, M. T. (2015). Arrhythmogenic calmodulin mutations affect the activation and termination of cardiac ryanodine receptor-mediated Ca^{2+} release. *Journal of Biological Chemistry*, *290*(43), 26151–26162.

Splawski, I., Timothy, K. W., Sharpe, L. M., Decher, N., Kumar, P., Bloise, R., … Keating, M. T. (2004). $Ca_V1.2$ calcium channel dysfunction causes a multisystem disorder including arrhythmia and autism. *Cell*, *119*(1), 19–31.

Stessman, H. A. F., Xiong, B., Coe, B. P., Wang, T., Hoekzema, K., Fenckova, M., … Eichler, E. E. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nature Genetics*, (January).

Stickel, F., Buch, S., Zoller, H., Hultcrantz, R., Gallati, S., Osterreicher, C., … Hampe, J. (2014). Evaluation of genome-wide loci of iron metabolism in hereditary hemochromatosis identifies PCSK7 as a host risk factor of liver cirrhosis. *Hum Mol Genet*, *23*(14), 3883–3890.

Studer, L., Vera, E., & Cornacchia, D. (2015). Programming and reprogramming cellular age in the era of induced pluripotency. *Cell Stem Cell*, *16*(6), 591–600.

Suh, W. (2017). A new era of disease modeling and drug discovery using induced pluripotent stem cells. *Archives of Pharmacal Research*, *40*(1), 1–12.

Suzuki, S., Ono, N., Furusawa, C., Ying, B.-W., & Yomo, T. (2011). Comparison of Sequence Reads Obtained from Three Next-Generation Sequencing Platforms. *PLoS ONE*, *6*(5), e19534.

Swanson, A. G. (1963). Congenital insensitivity to pain with anhydrosis: a unique syndrome in two male siblings. *Archives of Neurology*, *8*(3), 299–306.

Takahashi, K. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, *131*(5), 861–872.

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, *126*(4), 663–76.

Tatham, M. H., Plechanovová, A., Jaffray, E. G., Salmen, H., & Hay, R. T. (2013). Ube2W conjugates ubiquitin to α-amino groups of protein N-termini. *The Biochemical Journal*, *453*(1), 137–45.

Tennessen, J. a, Bigham, A. W., O'Connor, T. D., Fu, W., Kenny, E. E., Gravel, S., … Akey, J. M. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science*, *337*(6090), 64–9.

The HIPAA Privacy Rule. (2000). Retrieved April 11, 2017, from https://www.hhs.gov/hipaa/for-professionals/privacy/index.html?language=es.

Thevenon, J., Duffourd, Y., Lefebvre, M., Feillet, F., Steinmetz, A., Huet, F., & Chouchane, M.

(2016). Diagnostic odyssey in severe neurodevelopmental disorders : toward clinical whole-exome sequencing as a first-line diagnostic test. *Clinical Genetics*, *89*(6), 700–707.

Thiel, W. H., Chen, B., Hund, T. J., Koval, O. M., Purohit, A., Song, L.-S., … Anderson, M. E. (2008). Proarrhythmic defects in Timothy syndrome require calmodulin kinase II. *Circulation*, *118*(22), 2225–2234.

Thomson, J. A., Itskovitz-eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M. (1998). Embryonic stem cell lines derived from human blastocysts. *Science*, *282*(5391), 1145–1147.

Thornton-Wells, T. A., Moore, J. H., & Haines, J. L. (2004). Genetics, statistics and human disease: analytical retooling for complexity. *Trends in Genetics*, *20*(12), 640–647.

Triess, C., von Figura, G., Stuhrmann, M., Butzeck, B., Krayenbuehl, P. A., Strnad, P., & Kulaksiz, H. (2012). Diagnosis of hereditary hemochromatosis in the era of genetic testing. *Dig Dis Sci*, *57*(11), 2988–2994.

Trokovic, R., Weltner, J., Noisa, P., Raivio, T., & Otonkoski, T. (2015). Combined negative effect of donor age and time in culture on the reprogramming efficiency into induced pluripotent stem cells. *Stem Cell Research*, *15*(1), 254–262.

Ueno, S., Weidinger, G., Osugi, T., Kohn, A. D., Golob, J. L., Pabon, L., … Murry, C. E. (2007). Biphasic role for Wnt/beta-catenin signaling in cardiac specification in zebrafish and embryonic stem cells. *Proceedings of the National Academy of Sciences*, *104*(23), 9685–9690.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), e115.

Vacca, M., Della Ragione, F., Scalabrì, F., & D'Esposito, M. (2016). X inactivation and reactivation in X-linked diseases. *Seminars in Cell & Developmental Biology*, *56*, 78–87.

Valenti, L., Fracanzani, A. L., Rametta, R., Fraquelli, M., Soverini, G., Pelusi, S., … Fargion, S. (2012). Effect of the A736V TMPRSS6 polymorphism on the penetrance and clinical expression of hereditary hemochromatosis. *J Hepatol*, *57*(6), 1319–1325.

Van Damme, P., Evjenth, R., Foyn, H., Demeyer, K., De Bock, P.-J., Lillehaug, J. R., … Gevaert, K. (2011). Proteome-derived peptide libraries allow detailed analysis of the substrate specificities of N(alpha)-acetyltransferases and point to hNaa10p as the post-translational actin N(alpha)-acetyltransferase. *Molecular & Cellular Proteomics*, *10*(5), M110.004580.

Van Damme, P., Støve, S., Glomnes, N., Gevaert, K., & Arnesen, T. (2014). A *Saccharomyces cerevisiae* model reveals *in vivo* functional impairment of the Ogden syndrome N-terminal acetyltransferase NAA10 Ser37Pro mutant. *Molecular & Cellular Proteomics*, *13*(8), 2031–41.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, *30*(9), 418–426.

Vassalle, M. (1977). The relationship among cardiac pacemakers. Overdrive suppression. *Circulation Research*, *41*(3), 269–77.

Veerman, C. C., Kosmidis, G., Mummery, C. L., Casini, S., Verkerk, A. O., & Bellin, M. (2015). Immaturity of human stem-cell-derived cardiomyocytes in culture: fatal flaw or soluble problem? *Stem Cells and Development*, *24*(9), 1035–1052.

Viggiano, E., Picillo, E., Cirillo, A., & Politano, L. (2013). Comparison of X-chromosome inactivation in Duchenne muscle/myocardium-manifesting carriers, non-manifesting carriers and related daughters. *Clinical Genetics*, *84*(3), 265–70.

Voges, H. K., Mills, R. J., Elliott, D. A., Parton, R. G., Porrello, E. R., & Hudson, J. E. (2017). Development of a human cardiac organoid injury model reveals innate regenerative potential. *Development*, *144*(6), 1118–1127.

Wan, J. C. M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J. D., Caldas, C., … Rosenfeld, N. (2017). Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nature Reviews. Cancer*, *17*(4), 223–238.

Wang, G., McCain, M. L., Yang, L., He, A., Pasqualini, F. S., Agarwal, A., … Pu, W. T. (2014). Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nature Medicine*, *20*(6), 616–623.

Wang, H., He, Y., Li, Y.-H., Su, Z.-E., Li, B., Huang, H.-L., … Pan, J.-W. (2017). High-efficiency multiphoton boson sampling. *Nature Photonics*, doi:10.1038/nphoton.2017.63 [Published online 01 M.

Warburg, O. (1956). On the origin of cancer cells. *Science*, *123*(3191), 309–14.

Warr, A., Robert, C., Hume, D., Archibald, A., Deeb, N., & Watson, M. (2015). Exome sequencing: current and future perspectives. *G3 (Bethesda)*, *5*(8), 1543–50.

Webster, G., & Berul, C. (2013). An update on channelopathies from mechanisms to management. *Circulation*, *127*, 126–140.

Wei, Z., Wang, W., Hu, P., Lyon, G. J., & Hakonarson, H. (2011). SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. *Nucleic Acids Research*, *39*(19), e132.

Weldon, W. F. R. (1902). Mendel's laws of alternative inheritance in peas. *Biometrika*, *1*(2), 228–254.

Whiteway, M., & Szostak, J. W. (1985). The ARD1 gene of yeast functions in the switch between the mitotic cell cycle and alternative developmental pathways. *Cell*, *43*(2 Pt 1), 483–92.

Wibom, R., Lasorsa, F. M. F., Töhönen, V., Barbaro, M., Sterky, F. H., Kucinski, T., … Wedell, A. (2009). AGC1 deficiency associated with global cerebral hypomyelination. *New England Journal of Medicine*, *361*(5), 489–495.

Williams, B. C., Garrett-Engele, C. M., Li, Z., Williams, E. V., Rosenman, E. D., & Goldberg, M. L. (2003). Two putative acetyltransferases, San and Deco, are required for establishing sister chromatid cohesion in drosophila. *Current Biology*, *13*(23), 2025–2036.

Williams, J. C., Xu, J., Lu, Z., Klimas, A., Chen, X., Ambrosi, C. M., … Entcheva, E. (2013). Computational optogenetics: empirically-derived voltage- and light-sensitive channelrhodopsin-2 model. *PLoS Computational Biology*, *9*(9), 17–19.

Willig, L. K., Petrikin, J. E., Smith, L. D., Saunders, C. J., Thiffault, I., Miller, N. A., … Kingsmore, S. F. (2015). Whole-genome sequencing for identification of Mendelian disorders in critically ill infants: a retrospective analysis of diagnostic and clinical findings. *The Lancet Respiratory Medicine*, *3*(5), 377–387.

Wimplinger, I., Rauch, A., Orth, U., Schwarzer, U., Trautmann, U., & Kutsche, K. (2007). Mother and daughter with a terminal Xp deletion: implication of chromosomal mosaicism and X-inactivation in the high clinical variability of the microphthalmia with linear skin defects (MLS) syndrome. *European Journal of Medical Genetics*, *50*(6), 421–31.

Wolf, N. I., & van der Knaap, M. S. (2009). AGC1 deficiency and cerebral hypomyelination. *New England Journal of Medicine*, *361*(20), 1997–1998.

Wright, C. F., Fitzgerald, T. W., Jones, W. D., Clayton, S., McRae, J. F., van Kogelenberg, M., … DDD study. (2015). Genetic diagnosis of developmental disorders in the DDD study: a scalable analysis of genome-wide research data. *Lancet*, *385*(9975), 1305–14.

Wu, M., Neilson, A., Swift, A. L., Moran, R., Tamagnine, J., Parslow, D., … Ferrick, D. A. (2007). Multiparameter metabolic analysis reveals a close link between attenuated mitochondrial bioenergetic function and enhanced glycolysis dependency in human tumor cells. *American Journal of Physiology. Cell Physiology*, *292*(1), C125-36.

Xu, R.-H., Chen, X., Li, D. S., Li, R., Addicks, G. C., Glennon, C., … Thomson, J. A. (2002). BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nature Biotechnology*, *20*(12), 1261–1264.

Yang, C., Chapman, A. G., Kelsey, A. D., Minks, J., Cotton, A. M., & Brown, C. J. (2011). X-chromosome inactivation: molecular mechanisms from the human perspective. *Human Genetics*, *130*(2), 175–185.

Yang, L., Soonpaa, M. H., Adler, E. D., Roepke, T. K., Kattman, S. J., Kennedy, M., … Keller, G. M. (2008). Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature*, *453*(7194), 524–8.

Yang, X., Pabon, L., & Murry, C. E. (2014). Engineering adolescence: maturation of human pluripotent stem cell-derived cardiomyocytes. *Circulation Research*, *114*(3), 511–523.

Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., … Eng, C. M. (2013). Clinical whole-exome sequencing for the diagnosis of Mendelian disorders. *New England Journal of Medicine*, *369*(16), 1502–1511.

Yazawa, M., & Dolmetsch, R. E. (2013). Modeling Timothy syndrome with iPS cells. *Journal of*

Cardiovascular Translational Research, *6*(1), 1–9.

Yazawa, M., Hsueh, B., Jia, X., Pasca, A. M., Bernstein, J. A., Hallmayer, J., & Dolmetsch, R. E. (2011). Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome. *Nature*, *471*(7337), 230–4.

Yi, C. H., Pan, H., Seebacher, J., Jang, I.-H., Hyberts, S. G., Heffron, G. J., … Yuan, J. (2011). Metabolic regulation of protein N-alpha-acetylation by Bcl-xL promotes cell survival. *Cell*, *146*(4), 607–20.

Yoon, H., Kim, H.-L., Chun, Y.-S., Shin, D. H., Lee, K.-H., Shin, C. S., … Park, J.-W. (2014). NAA10 controls osteoblast differentiation and bone formation as a feedback regulator of Runx2. *Nature Communications*, *5*, 5176.

Young, W., D'Souza, S. L., Lemischka, I. R., & Schaniel, C. (2012). Patient-specific induced pluripotent stem cells as a platform for disease modeling, drug discovery and precision personalized medicine. *Journal of Stem Cell Research & Therapy*, (S10), 10.

Yu, J., Vodyanik, M. a, Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., … Thomson, J. a. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science*, *318*(5858), 1917–20.

Yu, X., & Sun, S. (2013). Comparing a few SNP calling algorithms using low-coverage sequencing data. *BMC Bioinformatics*, *14*(1), 274.

Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J. D., … Lifton, R. P. (2013). *De novo* mutations in histone-modifying genes in congenital heart disease. *Nature*, *498*(7453), 220–3.

Zeng, Y., Min, L., Han, Y., Meng, L., Liu, C., Xie, Y., … Shou, C. (2014). Inhibition of STAT5a by Naa10p contributes to decreased breast cancer metastasis. *Carcinogenesis*, *35*(10), 2244–53.

Zhang, G., & Nebert, D. W. (2017). Personalized medicine: genetic risk prediction of drug response. *Pharmacology & Therapeutics*, pii: S0163-7258(17)30050-5.

Zhang, J., Klos, M., Wilson, G. F., Herman, A. M., Lian, X., Raval, K. K., … Kamp, T. J. (2012). Extracellular matrix promotes highly efficient cardiac differentiation of human pluripotent stem cells: the matrix sandwich method. *Circulation Research*, *111*(9), 1125–1136.

Zhang, J., Wilson, G. F., Soerens, A. G., Koonce, C. H., Yu, J., Palecek, S. P., … Kamp, T. J. (2009). Functional cardiomyocytes derived from human induced pluripotent stem cells. *Circulation Research*, *104*(4), e30-41.

Zhang, J., Wilson, G. F., Soerens, A. G., Koonce, C. H., Yu, J., Palecek, S. P., … Kamp, T. J. (2009). Functional cardiomyocytes derived From human induced pluripotent stem cells. *Circulation Research*, *104*(4), e30–e41.

Zhang, P., Su, J., & Mende, U. (2012). Cross talk between cardiac myocytes and fibroblasts: from multiscale investigative approaches to mechanisms and functional consequences. *Am J*

Physiol Heart Circ Physiol, *303*(12), 1385–1396.

Zhu, R., Millrod, M. A., Zambidis, E. T., & Tung, L. (2016). Variability of action potentials within and among cardiac cell clusters derived from human embryonic stem cells. *Scientific Reports*, *6*, 18544.

Zipes, D. P., Camm, A. J., Borggrefe, M., Buxton, A. E., Chaitman, B., Fromer, M., … Mcgregor, K. (2006). *ACC/AHA/ESC 2006 guidelines for management of patients with ventricular arrhythmias and the prevention of sudden cardiac death--executive summary: a report of the American College of Cardiology/American Heart Association Task Force and the European Societ*. *Circulation* (Vol. 114).

Zook, J. M., Chapman, B., Wang, J., Mittelman, D., Hofmann, O., Hide, W., & Salit, M. (2014). Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nature Biotechnology*, *32*(3), 246–251.

Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A*, *109*(4), 1193–1198.

Zwaka, T. P., & Thomson, J. a. (2003). Homologous recombination in human embryonic stem cells. *Nature Biotechnology*, *21*(3), 319–21.

Zwi, L., Caspi, O., Arbel, G., Huber, I., Gepstein, A., Park, I.-H., & Gepstein, L. (2009). Cardiomyocyte differentiation of human induced pluripotent stem cells. *Circulation*, *120*(15), 1513–1523.