UNIVERSITY OF CALIFORNIA
SANTA CRUZ

Characterization of Huntington's Disease Cell Model In Gene Expression, Intracellular Glucose Levels And Genetic Variation

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR ENGINEERING AND BIOINFORMATICS

by

Gepoliano dos Santos Chaves

June 2020

The Dissertation of Gepoliano dos Santos Chaves
is approved:

Professor Nader Pourmand, Chair

Professor Mark Akeson

Professor Jeremy Sanford

Professor David Bernick

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Copyright © by

Gepoliano dos Santos Chaves

2020

Table of Contents

List of Figures

List of Tables

Abstract

Characterization of Huntington's Disease Cell Model in Gene Expression, Intracellular Glucose Levels and Genetic Variation

by

Gepoliano dos Santos Chaves

An increased risk of diabetes has been observed among Huntington's disease (HD) carriers, but there has not been follow-up investigation of the genetic nature of the risk. Here, we investigated diabetes phenotypes signatures in gene expression, DNA polymorphisms and intracellular glucose levels comparing rat and human data. Using a genome-wide association study (GWAS), RNA sequencing (RNA-Seq) analysis, and western blotting of *Rattus norvegicus* and human, we were able to identify a correlation between HD and DNA polymorphisms in the family of sortilins, tetraspanins and HLA/MHC proteins. We were also able to correlate SNP variants and gene expression with intracellular glucose levels using the nanopipette platform. Our results suggest that ST14A cells, from *R. norvegicus*, are a reliable model of HD, because they allowed confirmation of markers of diabetes, such as sortilins, glucose transporters and MHC proteins in our cell model. This model allowed experiments that highlight genes involved in mechanisms targeted by diabetes drugs, such as glucose transporters, and proteins controlling insulin release, related to mHTT. To the best of our knowledge, this is the first GWAS study using RNA-Seq data from both ST14A rat HD cell model and human HD. Similarly, for the

first time, glucose levels were detected at the single-cell level in a HD model using nanopipettes.

This work is dedicated to

The loving memory of my mother, my grandmother and all those whose lives were abbreviated due to the lack of appropriate care and education. I am confident that these two assets can grant some extension of life.

Acknowledgments

I would like to thank the members of my committee: Professor Mark Akeson, Professor Nader Pourmand, Professor Jeremy Sanford and Professor David Bernick, for their time and support. I am indebted to Dr. Nader Pourmand for his guidance, support and dedication provided to me and to the Pourmand group. He has taught me by example how to be a strong man in caring for the home family and the lab family. I am also indebted to my colleagues in the Pourmand Lab who were so kind, especially to Emrah Ozel, Akshar Lohith, Gonca Bulbul, Ramiz Alkassir, Joseph Olivier, John Stanley, Tonya Capillo, Garrett Graham and Carlos Leon. All other members of the Pourmand Lab: thank you! I want to give a special thanks students who joined our lab and whose eager to learn was so inspiring and encouraging. I also thank the instructors of courses I was a teaching assistant, which funded my research: Dr. Nader Pourmand, Dr. Rebecca DuBois and Dr. Peter Weiss. TAing was an amazing experience that expanded the limits of my cultural barriers. Finally, I acknowledge those who love me unconditionally and with whom my thoughts are everyday: my wife Cristiana and my son Victor. Thanks also to all my family members in Brazil and in the USA.

Funding: I acknowledge the Government of Brazil, Ministry of Education, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Science without Borders (SwB), for funding majority of my Ph.D. SwB has allowed the accomplishment of my dream of an international education. This work was also funded by the USA National Institutes of Health.

Chapter 1

Introduction

1.1 Biology and History of Huntington's Disease

Huntington's disease (HD) was first described in 1872 by George Huntington [10]. In the 1950s, Venezuelan physician Americo Negrette identified and described a disabling choreiform condition affecting individuals around region of Lake Maracaibo which conformed to the autosomal dominant mode of transmission of Huntington's Disease [1]. In 1972, the work lead by Negrette was presented in New York in a conference commemorating the centenary year of Huntington's Disease [1]. Present in the Conference was Nancy Wexler, an American psychoanalyst whose mother suffered from HD. Wexler organized a research team with collaborators in the USA and Venezuela and the group then started systematic studies of the disease. In collaboration with the group led by James Gusella from Harvard Medical School at Massachusetts General Hospital, Wexler participated in the identification by the method of linkage analysis, of the

genetic locus for HD in the short arm of chromosome 4 using linkage analysis. Their landmark paper was published in 1983 [11]. Further studies with the cohort of patients in Venezuela, also using linkage analysis, led to the precise identification and isolation of the pedigree and genomic sequence of the gene linked to Huntington's disease by the Huntington's Disease Collaborative Research Group. The gene, called huntingtin (HTT), is located around position 3M or 4p 16.3 in human chromosome 4, as described in 1993 [12, 1]. In Figure 1.1, I introduce those who have researched HD (including researchers whose pictures are not there) before me, because my work was only possible due to their research [1].

The history of the drama lived by the Wexlers and their scientific effort in establishing genetic consortia of HD carriers for the determination of a Huntington's disease pedigree, was described by Alice Wexler in the book Mapping Fate, published in 1996. Collaborative research on HD led to the discovery of the disease as a neurodegeneration caused by a CAG trinucleotide expansion in a region of chromosome 4. CAG expansion results in a gain of function mutation by the huntingtin protein, which alters the chemical properties of the protein, making it more insoluble and prone to aggregate with itself and other proteins (Figure 1.2).

Aggregated mutant huntingtin is toxic to neurons and is the main cause of the neurodegenerative process, leading to the death of neurons in the basal ganglia. The exact cellular role of the huntingtin protein is still not clear, and therefore a matter of debate on how exactly the mutant protein performs its toxic effects. Here, I describe huntingtin from the perspective of a scaffolding protein, working in close proximity with

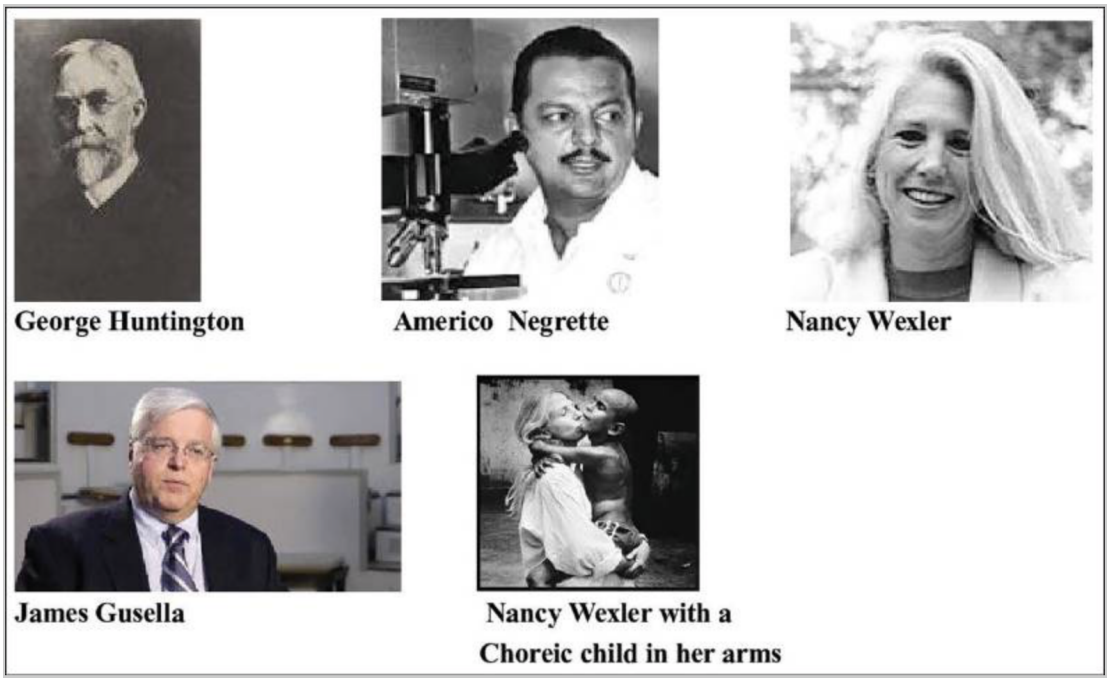

Figure 1.1: Scientists involved in early description and characterization of the pedigree of Huntington's disease. George Huntington described the condition in 1872. Americo Negrette discovered and diagnosed a cohort of HD individuals around Lake Maracaibo in Venezuela in the 1950s. The Venezuelan cohort was essential for identification of the HD pedigree and isolation of *HTT* gene sequence by Nancy Wexler, James Gusella and others in the 1980s and 1990s. Figure reprinted from [1] , with permission.

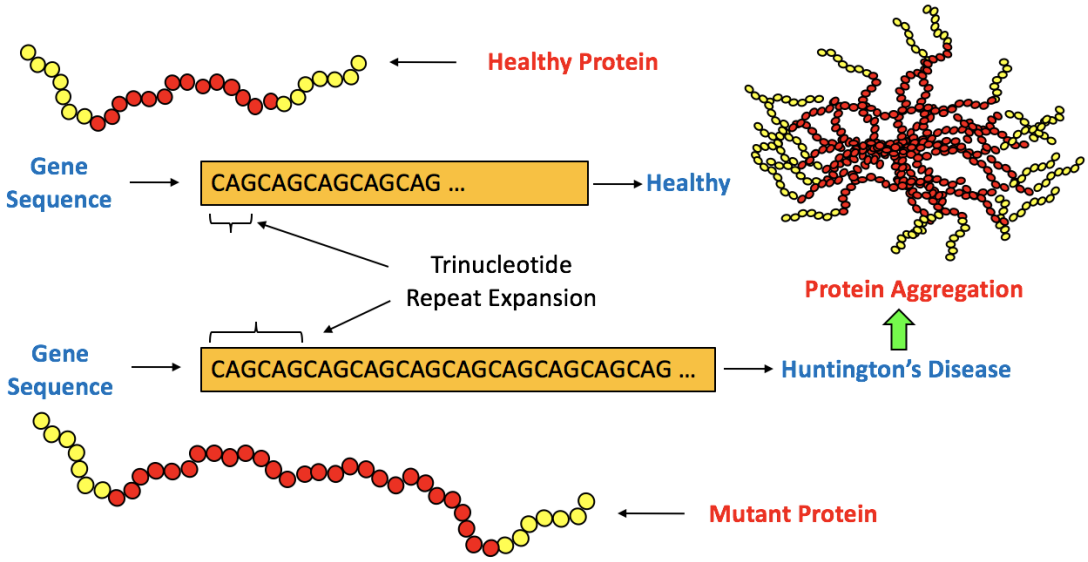

Figure 1.2: Huntington's disease CAG expansion and protein aggregation in affected individuals. Trinucleotide expansion in exon 1 of *HTT* leads to translation of a poly-glutamine (Q) trait in the mutant protein. This poly-Q trait leads to a gain of function by the protein, causing an increase in insolubility, protein aggregation and cytotoxicity.

molecular motor complex proteins such as Dynactin, Dynein and Kinesin-1 [2]. This function of HTT has therefore, important consequences for the intracellular traffic of vesicles using cytoskeleton proteins such as microtubules (Figure 1.3) [2].

The cellular toxic effects of the *mHTT* mutation is caused by and depends on the CAG trinucleotide expansion range in exon 1 of *HTT*. The CAG range in Huntington's disease is described as follows. Fewer than 27 CAGs in exon 1 of *HTT* is translated into healthy protein. Between 35 and 40 CAG repeats in exon 1 leads to the translation of a protein that does not lead to fully penetrant HD. Fully penetrant HD is determined by more than 40 CAG repeats in exon 1. Aggregation of huntingtin protein in fully penetrant HD leads to neuronal death and formation of empty spaces between the cerebral grooves (sulci) of the basal ganglia as disease progresses. Figure

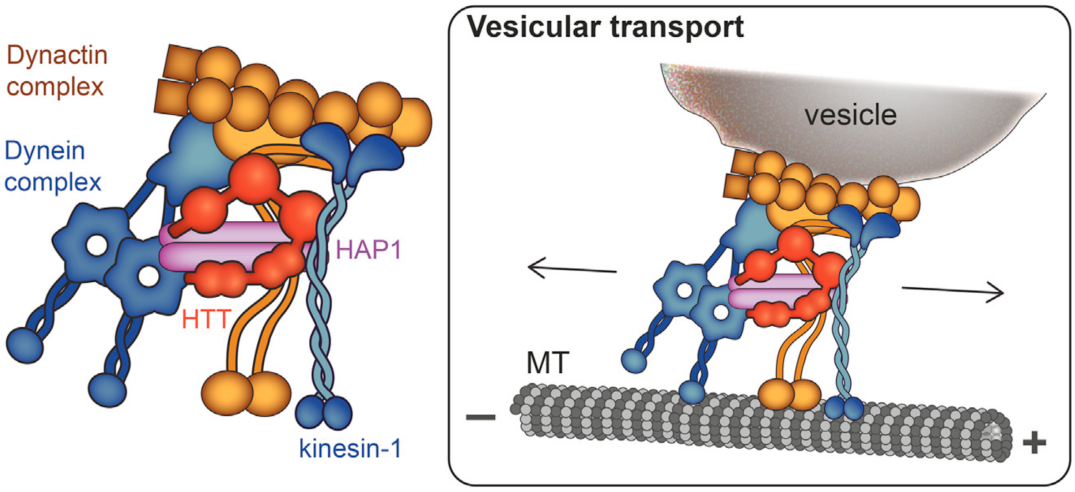

Figure 1.3: Evidence suggests a role for huntingtin as a scaffolding protein, acting in close contact with proteins of the molecular motor complex such as Dynactin, Dynein and Kinesin-1. This function is important for the traffic of cargo molecules inside vesicles transported using the cytoskeleton elements such as microtubules. Protein encoded by *HTT* locus aggregates and present problems for intracellular traffic. Figure reprinted from [2], with permission.

1.4 illustrates the presence of empty spaces caused by neuronal death in sulci present in the Basal Ganglia of HD carriers.

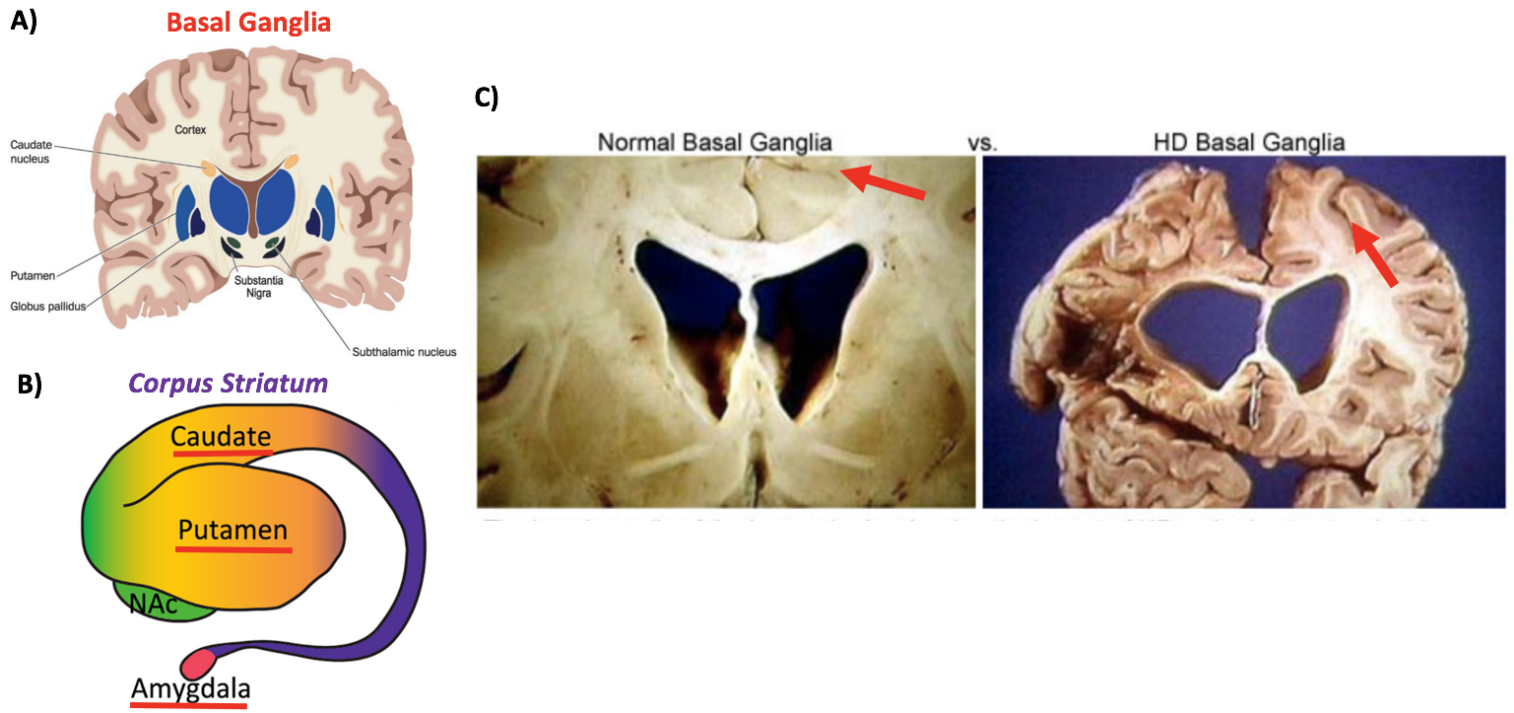

Figure 1.4: Effects of *HTT* mutation on the Basal Ganglia and Corpus Striatum of affected individuals. (A) Frontal plane visualization of the Basal Ganglia, located in the central portion of the brain. (B) Visualization of the striatum with closer indication (red underlines) of the three main component parts of the Corpus Striatum: Caudate nucleus, Putamen and Amygdala. (C) Comparison of *post-mortem* Basal Ganglia of normal and HD individuals (red arrows). Note the increased spacing between brain giri and sulci (cerebral grooves). Figure reprinted from [3, 4] and HOPES: Huntington's Outreach Project for Education at Stanford, with permission.

Protein aggregation and consequent neuronal death can be perceived at the body level of a HD carrier as degeneration progresses, in impaired movement, impaired cognitive function, development of psychiatric disorders and finally death. Death can occur as a normal consequence of disease progression or a direct impact in the psychological reward system leading to suicidal thoughts. One of the most sensitive regions of the brain and therefore one of the earliest affected organs by HD is the striatum, directly relating to the effects of HD in choreiform movements and emotional aspects of the HD individual. The striatum is a critical component of the physical motor and psychological reward systems [5]. The organ receives glutamatergic and dopaminergic inputs from different parts of the nervous system and serves as the primary input to the rest of the basal ganglia (Figure 1.5). Degeneration of the striatum in HD may start as early as 15 years of age [5] and is responsible for the high degree of degradation of the organ. Degradation of striatum impaires *post-mortem* sequencing of RNA molecules for evaluation of gene expression in this region, because aggregation of mHTT marks RNA molecules to destruction and/or degradation in neurons targeted to death (Figure 1.4C).

In order to overcome degeneration of the striatum and allow better characterization of effects of *HTT* in this organ, research groups have tried to isolate cell models from mouse striata. One example is the use of retroviral transduction of the SV40 Large T Antigen into cells from the embryonic *Rattus norvegicus* striatum primordia. These cells, called ST14A cells, were generated in Italy and acquired in our laboratory from the Coriell Institute [13]. ST14A cells conditionally immortalized and induced to stably

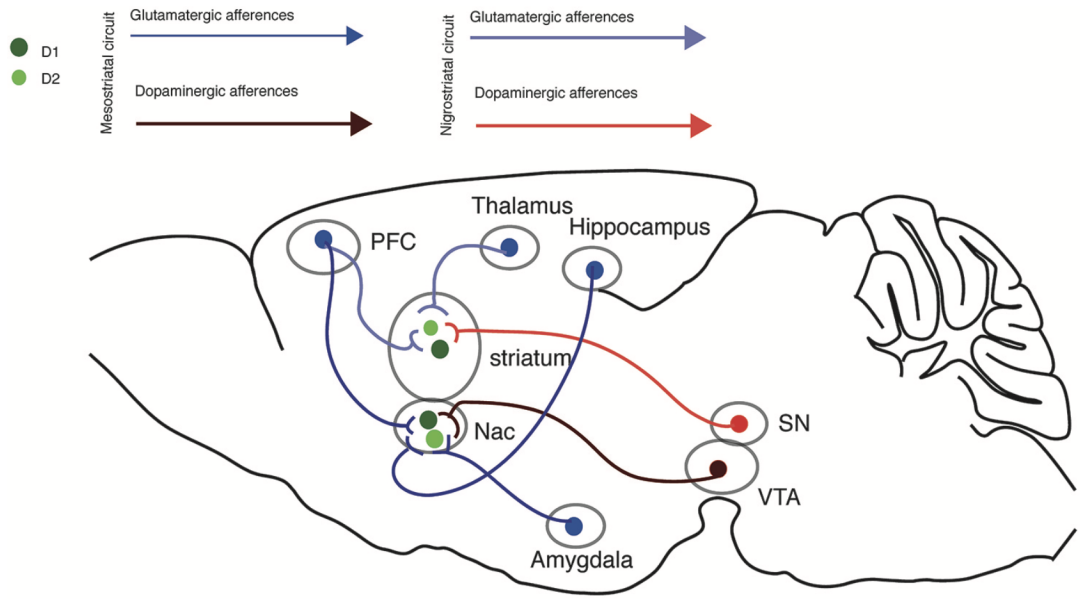

Figure 1.5: Participation of Corpus Striatum in the excitatory circuits connecting several parts of the brain. Dopaminergic and Glutaminergic circuits communicate distinct regions such as Thalamus, Hippocampus, Striatum, Nucleus Accumbens and Amygdala. These circuits compose the motor and cognitive function, as well as reward processing, and are impacted by aggregation of mutant huntingtin ($mHTT$) in individuals carrying HD. One important consequence of the impact of $mHTT$ in reward processing is an increased frequency of suicidal thoughts among HD individuals. PFC: prefrontal cortex; Nac: Nucleus accumbens; SN: Substantia Nigra; VTA: Ventral Tegmental Area. Figure reprinted from [5] with permission.

express 15 CAG repeats in the human huntingtin are called wild type cells, whereas cells immortalized and induced to stably express 120 CAG repeats in human mutant huntingtin protein are called mutant cells. As I discuss below, mHTT CAG expansion affected the expression of more than 3000 gene transcripts. Because the number of repeats in so dramatic in the mutant ST14A cells compared to the range that causes human Huntington's disease, we expected that it would dramatically affect Huntington's disease phenotypes. The described function of huntingtin in intracellular traffic (depicted in Figure 1.3) is critical for transmission of the nervous impulse in the Central Nervous System (CNS). Vesicles derived from cellular organelles such as endoplasmic reticulum and Golgi apparatus allow the necessary transport of neurotransmitters and molecules important in other cellular functions such as glucose receptor translocation. This role of huntingtin in subcellular traffic explains well an increase in risk of diabetes observed in HD patients. The explanation can be hypothesized as a model of pleiotropy of diabetes genes in HD. Pleiotropy is the biological phenomenon in which one genomic locus influences more than one trait or phenotype. Therefore, we hypothesized that pleiotropy of diabetes genes in Huntington's disease patients would be dependent on the biological function of the huntingtin protein explained in Figure 1.3. An increased risk of diabetes in Huntington's disease patients through mechanisms shown in Figure 1.3 raises the question of what the risk looks like in terms of Type 1 and Type 2 Diabetes phenotypes in Huntington's disease. Next, we explore the effects of $mHTT$ on the intracellular traffic of proteins or translocation of Glucose Receptor 4 (GLUT4), as pertains to T2D mechanisms.

1.2 Enhancement of Diabetes Phenotypes in *mHTT* Carriers

1.2.1 Type 2 Diabetes (T2D) Phenotypes

We hypothesized that pleiotropy of diabetes genes was a phenomenon enhanced by the presence of *mHTT* in HD carriers. The effect of *mHTT* in diabetes phenotypes can be investigated from the standpoint of the model of translocation of Glucose Receptor 4 to the plasma membrane for uptake of glucose, well characterized in diabetes models and depicted in Figure 1.6. In this model, binding of insulin to its cellular receptor (1) leads to transduction of the insulin signal. In the case we illustrate here, the insulin effect is the transport of GLUT4 to the cellular membrane, for uptake of glucose (2). In HD carriers, however, this process is affected due to the impairment of proper intracellular traffic caused by the presence of mHTT acting as the scaffolding protein in the molecular motor complex (3 in Figure 1.6; Figure 1.3). Therefore, we hypothesize that, in HD carriers, the ability to translocate GLUT4 to the cell membrane (Figure 1.6, 2) is impaired by mHTT (Figure 1.6, 3). I hypothesize that this phenomenon, Type 2 diabetes vesicle traffic impairment (also referred to as **insulin sensitivity**), in which diabetics loose the ability to clear blood glucose levels properly because of problems in the translocation of GLUT4, is worsened by mHTT. T2D vesicle traffic impairment, well documented in diabetes, also involves a class of genes called sortilins [6]. When explaining Type 1 Diabetes (T1D) phenotypes, I will discuss how mHTT can also affect subcellular traffic of immunological proteins.

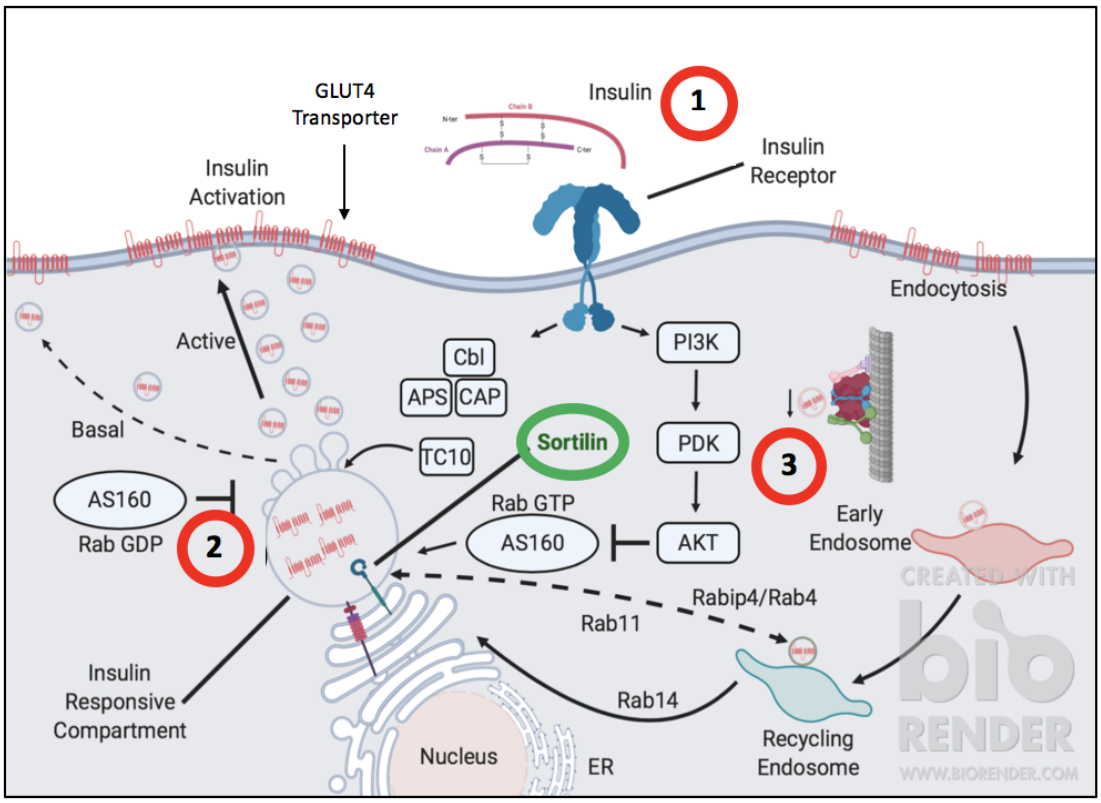

Figure 1.6: Subcellular traffic impairment in HD carriers due to mHTT in the translocation of GLUT4 receptor protein. Insulin binding to its cellular receptor (1) and consequent translocation of GLUT4 to the plasma membrane (2) is prevented from working properly by the presence of mHTT, interacting with the molecular motor complex proteins (3). Figure modified from [6].

Insulin and GLUT4 proteins are integral markers of diabetes as shown in the transduction mechanism depicted in Figure 1.6. Another family of genes involved in this process is that of sortilins. Sortilins have an intuitive name for their role in sorting of GLUT4 receptor and are involved in diabetes and Alzheimer's disease [6, 14, 8, 15]. Because the major marker of T2D is the blood levels of glucose, which are affected by the mechanism depicted in Figure 1.6, we became interested in measuring glucose levels in our Huntington's disease ST14A cell model. Previously, our group had reported on the development of a glucose sensor using the nanopipette platform [16]. In Chapter 2, I report on the use of nanopipettes to measure single cell glucose levels. In this project, I cultivated the ST14A cells with different glucose and insulin levels and prepared the cells for direct nanopipette sensing. I also performed the RNA-Seq analysis and the molecular biology characterization of different genes using western blot, PCR and quantitative PCR techniques. Below, I discuss how nanopipettes are employed to measure intracellular glucose levels.

1.2.2 Glucose Sensing Using Nanopipettes

Nanopipettes are small glass needles fabricated with tip diameters ranging from 50 to 100 nm [9]. The small diameter of the nanopipette tip allows the device to repeatedly interrogate a single cell without damage. An electrochemical system allows nanopipettes to perform various functions and assessments in the cell, including injection, aspiration, and real-time sensing [17]. For instance, I have introduced controlled amounts of plasmid into specific regions in a femtoliter scale (Injection) to observe

protein expression. I followed the injection of fluorescent dyes to show that multiple interrogation of single cells can be performed without damage to the cell; this work is covered in the review paper published by our group (Appendix A). I have also isolated and removed specific subcomponents of a living cell (e.g. mitochondria, nuclei or RNA molecules) (Aspiration). These results as well as the RNA-Seq analysis necessary for the study of mRNA molecules isolated from the nuclei of several cell types, are also discussed in the review paper in Appendix A. Finally, nanopipettes allow real time measurement of glucose, pH, proteins and other analytes in a living cell (Sensing). This measurement is accomplished without killing the living cell or interfering with intracellular glucose concentrations as occurs with traditional cell lysing methods. With these applications, nanopipettes present a great potential for breakthrough discoveries in biotechnology because the applications accommodate genomic and sensing analysis and comparison between these two distinct fields (Figure 1.7). Single-cell glucose level results are discussed in Chapter 2, where we used nanopipettes to measure intracellular levels of glucose in the ST14A HD cell model. For a complete picture of the principle of glucose detection using nanopipette, the reader is referred to the original report [16]. As I mentioned above, sortilins are not involved solely in T2D through the translocation of GLUT4 receptors to the cellular membrane for glucose uptake. In section 1.2.3, I discuss the involvement of sortilins in Type 1 Diabetes phenotypes that help explain an increased risk of diabetes in HD carriers.

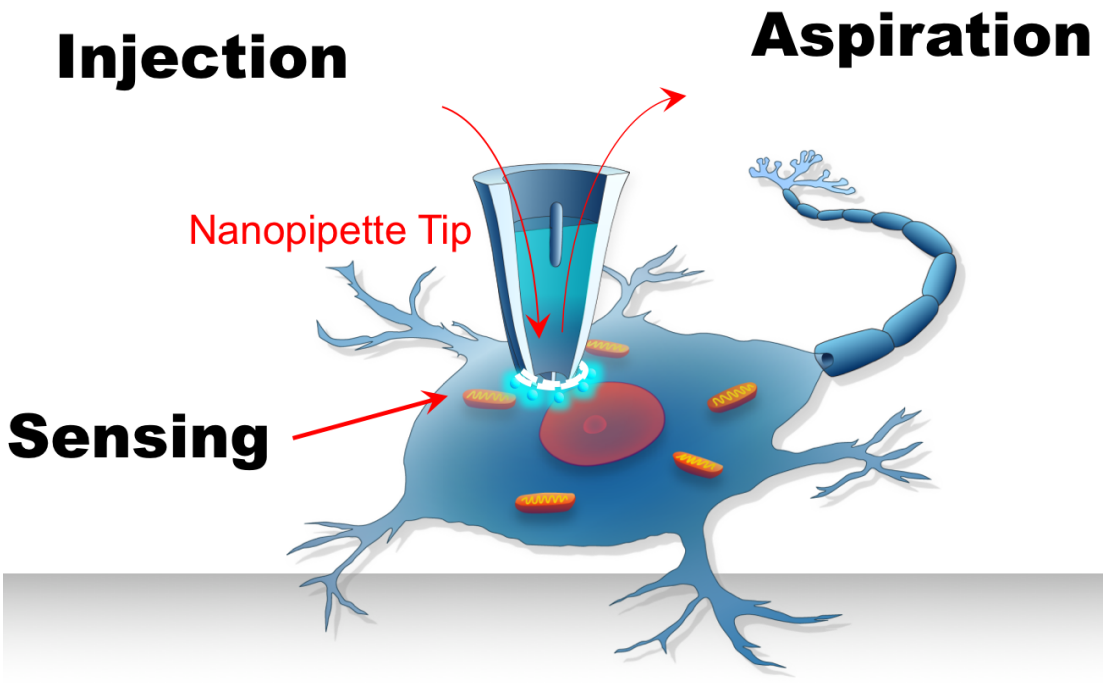

Figure 1.7: The nanopipette platform allows longitudinal interrogation of single cells and controlled introduction of nano-material (Injection), removal of cellular components (Aspiration) and measurement of intracellular metabolites (Sensing). Here, our group used nanopipettes to measure intracellular levels of glucose in ST14A cells.

1.2.3 Type 1 Diabetes (T1D) Phenotypes

1.2.3.1 Translocation of Immunological Vesicles to the Plasma Membrane is Problematic in Diabetes and HD

In subsection 1.2.1, Type 2 Diabetes (T2D) Phenotypes, I discussed the role of mHTT in impairing translocation of GLUT4 protein to the cellular membrane and consequent effects for insulin sensitivity and scavenging of blood glucose levels (Figure 1.6). This phenomenon is well described in adipocytes and muscle cells, but, in general, does not consider the effect of HTT or mHTT in vesicular intracellular translocation and/or transport, and the consequent implications for *mHTT* carriers. To show that a more complete picture of the molecular events involved in diabetes phenotypes observed in HD is provided by the notion of impairment in the intracellular traffic promoted by mHTT, I describe now how translocation of proteins of two important immunological groups, MHC and tetraspanins, can be impacted in their traffic from the endoplasmic reticulum and Golgi apparatus to the cellular membrane. The importance of these two groups of proteins, which can also be understood as genomic loci, resides in their ties to T1D. MHC is the main genomic locus associated with T1D and other autoimmune diseases, due to its involvement in presentation of antigens in the cellular membrane of antigen-presenting cells (APCs) to the T-cell receptor [18, 19]. Not only tetraspanins and MHC, but also, sortilins, are actively involved in processing antigens displayed by MHC proteins in the cellular membrane. In order to observe the hypothetical role of mHTT in intracellular traffic of immunological proteins, the reader is referred to Figure

1.8. That schematic aims to show how the presence of mHTT (1) affects endocytosis of vesicles containing microbes that are endocytosed using cytoskeleton and molecular motor complex proteins (2). Although proteins derived from extracellular microbes are processed via class II MHC (3) and tumor or viral cytoplasmic antigens are processed via class I MHC (4), both class I and class II MHC antigens undergo processing via endoplasmic reticulum (5) and Golgi apparatus (6). Because of the similarity between the mechanisms depicted in Figure 1.8 and the mechanism of insulin sensitivity in T2D (Figure 1.6), I hypothesized that the immunological genes associated with T1D should also be active in diabetes phenotypes of Huntington's disease. Validating this idea and establishing a parallel between the T2D center (muscle and adipocytes) and the T1D center (pancreas), Kebede *et al.* (2014) demonstrated that sortilin SORCS1 is a diabetes marker with a role in translocation (also understood as transport or secretion) of insulin in pancreatic β-cells to the blood-stream [20]. This work provides insight because it dissociates sortilins from the idea of T2D in the context of GLUT4 translocation and insulin sensitivity in muscles and adipocytes, and includes sortilins more in the context of T1D in the pancreatic β-cells depicted in Figure 1.8. In pancreatic β-cells, insulin is translocated to the exterior of the cell by secretion via endoplasmic reticulum and Golgi apparatus. Because sortilins are not restricted to T2D phenotypes, but instead also influence TD1 (which is quite surprising from the standpoint of the dichotomy T1D vs. T2D), I next started to review and revise the known roles of sortilins in T1D and events of T1D such as phagosome maturation and antigen processing.

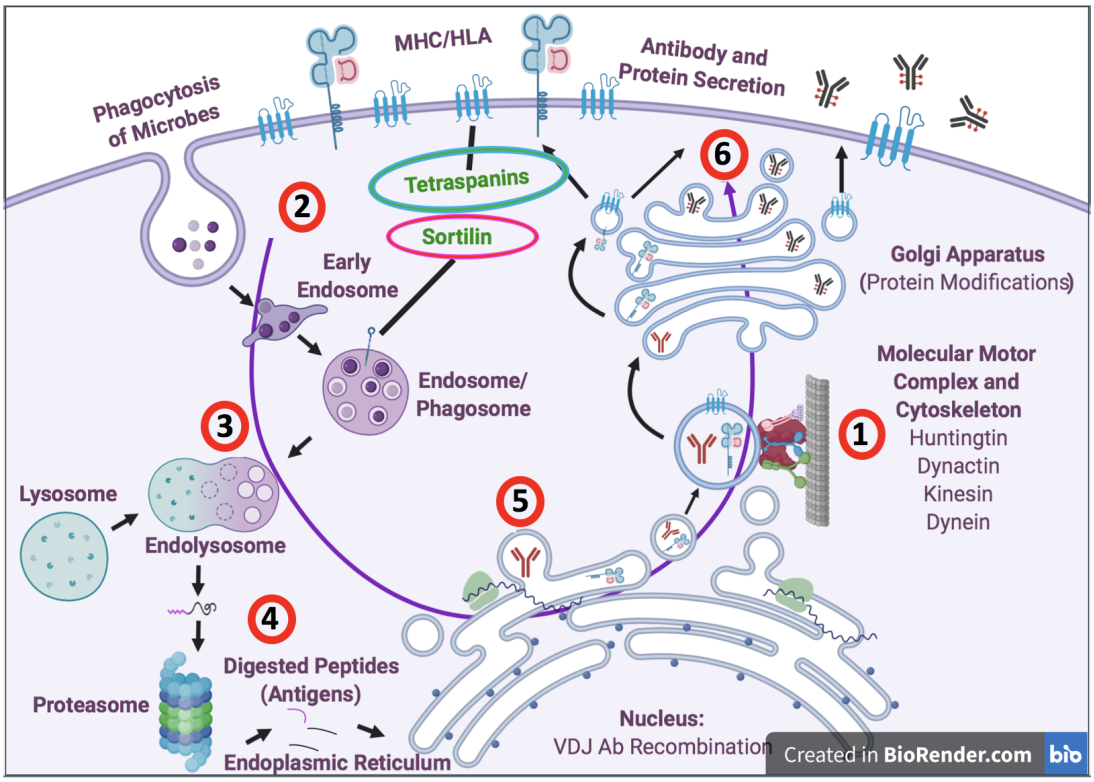

Figure 1.8: Translocation of immunological proteins to the cellular membrane is affected by mHTT. As exemplified in insulin sensitivity mechanisms, mHTT impacts translocation of vesicles containing immunological proteins such as MHC, tetraspanins and antibodies (1). Endocytosis of extracellular microbes (2) and their processing in the endosomes (3) follow class II MHC processing. Cytoplasmic proteins are processed by proteasome (4) and undergo class I MHC processing. Both classes I and II MHC-bound antigens are translocated to the plasma membrane via endoplasmid reticulum (5) and Golgi apparatus (6). Literature reference for these mechanisms can be found along the text. The end result of this mechanims is that in HD, GLUT4, tetraspanins, insulin, antibodies and MHC proteins reach the cellular membrane or exterior of the cell via mHTT-impacted cytoskeleton transport.

1.2.3.2 Sortilins and Tetraspanins as Markers of T1D

In general, the scientific literature does not try to differentiate the roles of sortilins as either a T1D or T2D marker. Only by considering insulin sensitivity as an inability of a diabetic individual to translocate GLUT4 to the cellular membrane, and by factoring in the known role for sortilins in that process (Figure 1.6), could one argue this to be a sortilin involment in T2D. However, if we consider that T1D is primarily a deficiency of the pancreas to produce or release insulin either because of autoimmune destruction of pancreatic cells or failure of these cells to secrete insulin, we realize that the work of Kebede *et al.* connects T2D and T1D by identifying sortilins as diabetes markers in preventing insulin release in the pancreas [20]. Surprisingly, therefore, sortilins not only have a role in insulin sensitivity or T2D, by directly working on translocation of GLUT4 in muscles and adipocytes (Figure 1.6), but sortilins also participate in the release of insulin from the pancreas (T1D, Figure 1.8) [20]. Two other literature papers need to be mentioned to further illustrate the role of sortilins in immunological processes. These papers mention the role of sortilins in subcellular transport of antigens derived from phagosomes and should be considered in the analysis of Figure 1.8. This process is referred to as phagosome maturation. In the process of phagosome maturation, antigens are processed via class II MHC, because phagosomes derive from endocytosis of extracellular proteins, and not cytoplasmic (viral and tumoral) proteins (Figure 1.8, parts 1 and 2). In the first paper, Vázquez (2016) characterized the role of sortilins in the maturation of phagosomes in macrophages [21]. In this study, phago-

cytosis of *Mycobacterium tuberculosis* is followed by sorting of proteins into endosomes for enzymatic digestion of parasite proteins, therefore originating antigens derived from the pathogen (Figure 1.8, parts 2 and 3). A second line of evidence, revised by Talbot (2019), suggests that in a special class of macrophages named alternatively associated macrophages (AAMs), antigen interaction with sortilins leads to degradation of antigens instead of display of antigens by class II MHC proteins in the cellular membrane [22]. Talbot highlights that this mechanism is important in organ transplantation. In both the Vázquez and Talbot reports, sortilins are involved in the processing of peptides originated by digestion of proteins to be presented in the cellular membrane by class II MHC proteins. As mentioned, class II MHC proteins bind antigens that are generated by proteolisis of extracellular proteins via endosomes and lysosomes (Figure 1.8, parts 2 and 3). This MHC processing mechanism can be contrasted with class I MHC processing in that class I MHC processing expresses viral and tumoral (cytosolic) antigens, processed by proteasome, in the cell membrane (Figure 1.8, part 4). Class II MHC proteins present antigens to CD4+ T cells [23]. The translocation of antigen-MHC to the cellular membrane happens by means of vesicle migration to the membrane using citoskeleton transport by molecular motor complex proteins. Therefore, translocation of vesicles containing MHC-peptide molecules for MHC restriction in the cellular membrane use mechanisms similar to translocation of GLUT4. Display to T-cell receptor and activation of the adaptive immune response is also known as MHC peptide restriction. I hypothesize that mHTT affects proper traslocation of the immunological MHC complexes for MHC peptide restriction in $mHTT$ carriers (Figure 1.9).

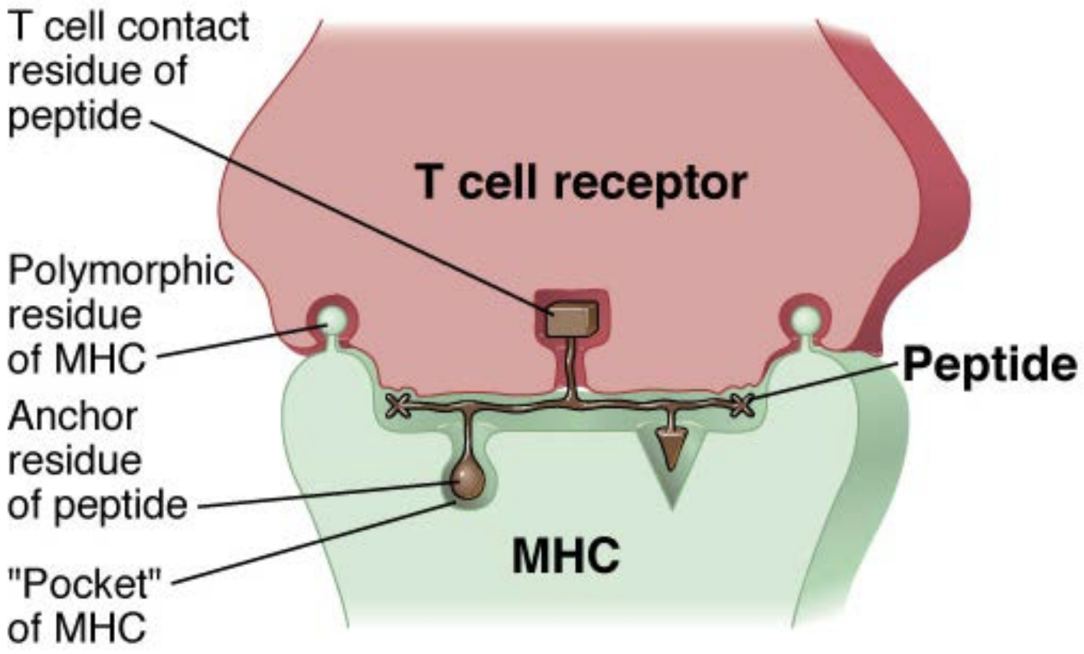

Figure 1.9: T-cell recognition of peptide-MHC complex, or MHC restriction. An MHC molecule binds and displays a peptide to the T-cell receptor. The T-cell receptor recognizes three important regions in the MHC: two polymorphic peptide domains in the MHC sequence and one domain in the antigenic peptide sequence. The antigenic peptide shows high affinity to amino-acids present in the MHC pocket. I hypothesize that in *mHTT* carriers, MHC restriction is impacted by the effect of mHHT in the translocation of vesicles containing peptide-MHC complex to the cellular membrane. Figure reprinted from [7] with permission.

1.2.3.3 MHC Peptide Restriction and Organization of the MHC Locus

MHC restriction is an early event in the adaptive immune response, and plays a role in adaptive immune response against self and non-self peptides. In normal conditions, adaptive immune responses target viral, bacterial or tumor peptides. In abnormal conditions, for example when there is underlying autoimmune disease, self-peptides are the target of adaptive immunity. Type 1 diabetes is an example of an autoimmune disease. Professional antigen-presenting cells (APCs), such as Dendritic Cells (DCs) are cells in charge of patrolling the immune system, sampling various types of proteins and determining whether the protein deserves destruction. In this way, DCs isolate samples from their environment, process these antigens (such as those derived from bacteria), and initiate adaptive immunity by activating CD4+ and CD8+ T-cells. In HD, due to the well characterized CAG expansion mutation in huntingtin exon 1, other important genomic regions such as the genomic locations encoding immunologically-related genes are overlooked in the role they play in the onset or progression of the disease. Considering the epigenetic effects of environmental factors such as physical exercise, education levels and food intake on symptoms of Huntington's Disease, potentially the progression or even the age of onset of disease can be retarded, or actions taken by the patient regarding these factors so that age of onset or progression of the disease is retarded. Given their involvement in processing and presentation of protein antigens to the immune-system, the genes encoded by the major histocompatibility complex locus, in humans named *HLA* (Human Leukocyte Antigen), are potentially important in

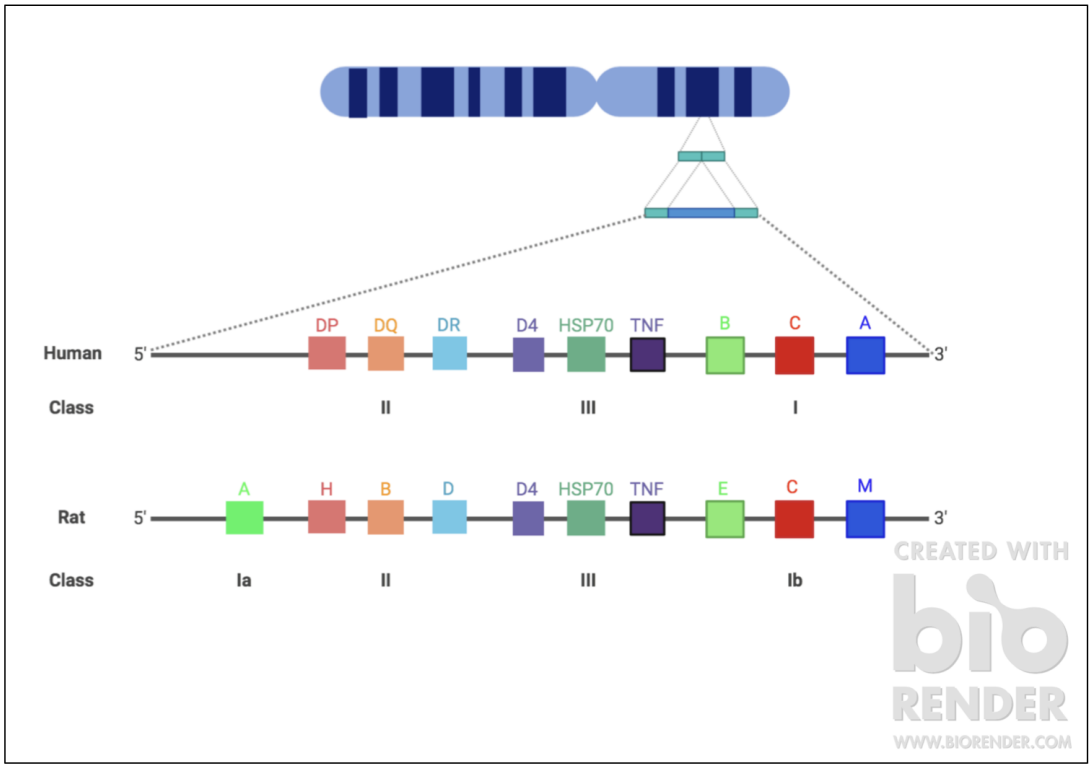

Figure 1.10: Representation of the general genomic organization of the major histocompatibility complex (MHC) locus in human and rat chromosomes. MHC genes locate in chromosome 6 in human and chromosome 20 in rat. In human, the MHC locus is called *HLA* (Human Leukocyte Antigen).

HD pathology. Much evidence suggests an immune system involvement in HD, either because MHC proteins participate in the pathogenesis of Type I diabetes, in which insulin-producing cells in the pancreas are destroyed, or because an immune response is mounted against mutant huntingtin [19, 18, 24, 25]. MHC genes were thought to play a role in HD pathology before the huntingtin locus was characterized [26], and current research supports this notion by suggesting MHC gene expression associated with HD [25]. Figure 1.10 depicts the genomic location of the main MHC genes. The figure also shows the different locations of the classes I, II, and III MHC genes.

MHC mRNAs are translated by ribosomes in the rough ER and translocated to the ER lumen where they are combined with antigens capable of triggering an immune response (Figure 1.8 part 5). These antigens can reach the ER lumen directly from proteasome degradation (Figure 1.8 part 4), endosome digestion (Figure 1.8 part 3), or can be acquired from phagocytosis of extracellular proteins (Figure 1.8 part 2). In the context of this dissertation, I hypothesize that the expression of mutant huntingtin protein has major consequences for MHC allele expression and peptide loading via class I and class II antigen processing. In both class I and class II MHC-peptide loading, the ER is the cellular organelle central to MHC protein synthesis. In Chapter 3, I further explore the concept that sortilins have an immunological function, as do MHC and tetraspanins [27]. Results of the RNA-Seq analysis and single cell glucose levels are discussed in the next section, Chapter 2, which were published as a research paper, shown in reference [28].

Chapter 2

Metabolic and transcriptomic analysis of Huntington's disease model reveal changes in intracellular glucose levels and related genes

Overview

The following is a transcript of our published paper, Chaves *et al.*, (2017) [28]. For this project, I cultivated the ST14A cells with different glucose and insulin levels and prepared the cells for direct nanopipette sensing. I also performed the RNA-Seq analysis and the molecular biology characterization of different genes using western blot, PCR and quantitative PCR techniques. These data represent a comprehensive integration of metabolic and transcriptomic expression data in rat HD cells, and help to identify

genes that may be involved in crosstalk between HD and diabetes.

Abstract

Huntington's Disease (HD) is a neurodegenerative disorder caused by an expansion in a CAG-tri-nucleotide repeat that introduces a poly-glutamine stretch into the huntingtin protein (mHTT). Mutant huntingtin (mHTT) has been associated with several phenotypes including mood disorders and depression. Additionally, HD patients are known to be more susceptible to type II diabetes mellitus (T2DM), and HD mice model develops diabetes. However, the mechanism and pathways that link Huntington's disease and diabetes have not been well established. Understanding the underlying mechanisms can reveal potential targets for drug development in HD. In this study, we investigated the transcriptome of mHTT cell populations alongside intracellular glucose measurements using a functionalized nanopipette. Several genes related to glucose uptake and glucose homeostasis are affected. We observed changes in intracellular glucose concentrations and identified altered transcript levels of certain genes including Sorcs1, Hh-II and Vldlr. Our data suggest that these can be used as markers for HD progression. Sorcs1 may not only have a role in glucose metabolism and trafficking but also in glutamatergic pathways affecting trafficking of synaptic components.

2.1 Introduction

Huntington's Disease (HD) is a progressive, autosomal dominant neurodegenerative disease that produces physical, mental and emotional changes due to loss of critically important brain neurons. The genetic basis for HD is an expansion of cytosine-adenine-guanine (CAG) repeats in the huntingtin (HTT or IT15) gene that leads to the formation of a prolonged polyglutamine (polyQ) tract in the N-terminal region of the mutant huntingtin protein (mHTT). The wild type huntingtin protein (HTT) is important for the intracellular transport and trafficking of proteins, organelles and vesicles. The expansion of glutamine (>36 repeats) produces a gain of function mHTT that affects the healthy function of cellular machinery, ultimately resulting in neurotoxicity and detrimental cell lethality in the brain [29]. Considerable research efforts have been made to elucidate the molecular and cellular mechanisms underlying HD pathology. The proposed mechanisms through which mHTT causes neurodegeneration include mutant protein aggregation, vesicle association, elevated oxidative stress, excitotoxicity, mitochondrial and transcriptional dysregulation [30, 31]. Evidence for several of these mechanisms has been found in post-symptomatic disease models or post-mortem brain samples. Controversially, during the pre-symptomatic stage of HD, cellular architecture and morphology have been found to be disrupted but neurodegeneration has been found to be minimal or absent [32, 33]. These studies have been limited to immunohistochemistry, fluorescence staining and immunoblot analysis, and provided static information of HD [34, 35]. However, little is known at the early stages of the disease about the

dynamics of metabolic and transcriptomic changes, which could be instrumental not only for HD therapy for diagnosed patients but also for at-risk individuals. Striatum, a subcortical part of the forebrain, is the most vulnerable part of the brain in HD, almost disappearing during the course of disease. One of the major clinical symptoms of HD is the loss of the medium spiny neurons in the striatum [36]. Therefore, although genome-wide gene expression studies have been conducted on targeted affected tissues in the post mortem human brain, it has proven difficult to detect specific changes in the human striatum [37]. To expand our understanding of changes in the transcriptional landscape of the affected cells, mRNA expression profiling was performed on cells isolated from the striatum of rat embryos expressing wild type and mutant HTT. As shown by Lin and Beal, detecting changes in carbohydrates can be instrumental for the development of useful therapies that slow down HD progression and reduce its severity at the early stages of the disease [38]. Recent evidence suggests that, similar to other neurodegenerative diseases, mice exhibit increased deregulation in brain energy metabolism at the late stages of HD [39]. A super-family of glucose transporter genes, the GLUT gene family, is responsible for the transport and uptake of glucose [40, 41]. The GLUT family includes 12 genes, each encoding one GLUT protein. Various members of this gene family were detected in the brain, although glucose transporter protein type I exclusively mediates transport across the blood-brain barrier [42, 43]. GLUT1, initially thought to be the only glucose transporter protein expressed in the brain, supports the basal metabolic needs of proliferating cells [44]. GLUT2 is primarily GLUT protein expressed in the pancreatic β-cells, liver and kidneys. In the pancreas, GLUT2 is believed

to act in conjunction with glucokinase in the glucose-sensing mechanism. GLUT4 is the insulin-responsive receptor and is primarily expressed in the heart, adipose tissue, and skeletal muscle, where it is responsible for the reduction in the glucose levels in the post-prandial rise of glycaemia. Insulin stimulates the translocation of intracellular vesicles containing GLUT4 to the plasma membrane, resulting in a 10- to 20-fold increase in glucose transport [45]. Current assessment methods for intracellular glucose levels in HD models are limited and performed with a combination of techniques including gas chromatography-mass spectrometry (GC-MS), liquid chromatography-mass spectrometry (LC-MS), uptake assays, and electrophysiology [46]. The invasiveness, low sensitivity, and specificity of these techniques prevent longitudinal study of biological model cells. Other major limitations of these techniques are the complex sample preparation methods during which cellular concentration of glucose can be altered. Therefore, accurate real time detection of glucose starting from the early and going to the later stages of the disease may provide biochemical evidence that can inform the design of novel drugs and therapeutic regimens. We recently reported the use of nanopipette technology for intracellular measurements [16, 47]. Importantly, because nanosensors are minimally invasive and not destructive to the cell, it is possible to take repeated measurements of the same cell. In this work, we employed glucose nanosensors for intracellular glucose measurements in HD cell models (ST14A) at different time points to investigate glucose metabolism over time. In parallel, we conducted transcriptomic analysis to understand temporal changes in RNA expression in model HD cells. We report impaired glucose metabolism in single rat striatum mHTT cells using our non-

destructive nano-glucose sensor. This is the first report determining the intracellular glucose levels in mHTT cells to be in the range of 0.8–1.5 mM. A 2.5 fold decrease in intracellular glucose levels in HD cells indicated an altered glucose metabolism that was further investigated by transcriptomic study. Significant differences were seen in glucose uptake and glucose homeostasis genes. Interestingly, there was an increased expression of Sorcs1 in mHTT cells which has been previously associated with late onset of Alzheimer's disease, another neurodegenerative disorder. Overexpression of Sorcs1 was confirmed by performing RT-qPCR which corroborates Sorcs1 can be used as a biomarker for HD progression.

2.2 Materials and methods

2.2.1 Reagents

Glucose Oxidase (GOx) from Aspergillus niger ≥100,000 units/g (Type VII, lyophilized, EC Number 232-601-0), ferrocene (98%), poly-l-lysine (PLL) (0.1% solution in water), glutaraldehyde (GA) (25% in water) and D-glucose were purchased from Sigma Aldrich (St. Louis, MO). Silver wires (125 µm) were supplied from A-M Systems (Sequim, WA). Glucose free Dulbecco's Modified Eagle Medium (DMEM), trypsin (0.25%, phenol red) and penicillinstreptomycin were bought from Gibco while Hank's Balanced Salt Solutions and fetal bovine serum were purchased from GE Healthcare (GE Healthcare). 500 µm gridded plates were obtained from Ibidi (Ibidi USA, Inc., Madison, WI).

2.2.2 Glucose nanosensor fabrication

Glucose nanosensors with an outside diameter of 1.00 mm and an inside diameter of 0.70 mm were fabricated from quartz capillaries (Sutter Instrument, Novato, CA) using a P-2000 laser puller (Sutter Instrument, Novato, CA) as described elsewhere [16]. Briefly, to create glucose nanosensors, the nanopipettes were modified as follows: Pulled nanopipettes were backfilled with a 15 µl solution containing 10 mM Ferrocene prepared in 100 mM PBS (pH 7, supplemented with 0.1 M KCl). An Ag/AgCl electrode was placed into the nanopipette as a working electrode while another Ag/AgCl electrode was immersed in the cell media as a reference/counter electrode. Nanopipettes were modified with PLL by backfilling the nanopipette interior. Then nanopipette surface was treated with a 10% (v/v) solution of glutaraldehyde (GA) for 30 min. GOx (900 mU) was then reacted with the activated nanopipette walls. All glucose nanosensors were calibrated in DMEM with standard glucose concentrations.

2.2.3 Cell culture

The Huntington's disease model cell lines that were used in this work are both striatal cell lines derived from rat embryos and obtained from Coriell Institute cell repository. The two cell lines express human HTT fragments of 15 and 120 polyglutamine repeats representing wild type (ST14A-Q15) and mHTT disease model (ST14A-Q120), respectively. All cells were cultivated and maintained in DMEM (supplemented with 10% fetal bovine serum, antibiotics) with various glucose concentrations. DMEM was supplemented with 1 nM insulin when needed to study effects of insulin.

2.2.4 Intracellular glucose measurement

The intracellular glucose measurement setup consists of an inverted microscope Olympus IX 70 with Spot Insight CMOS camera to image cells. The nanosensors are fixed to a microscope by a pipette holder (Axon Instruments). The holder is connected to an Axopatch 700B amplifier (Molecular Devices) for current measurement, an MP-285 micromanipulator (Sutter Instrument) for coarse control of the nanopipette positioning in the X, Y, and Z directions, a Nanocube piezo actuator (Physik Instrument) for fine control in the X, Y, and Z directions, and a PCIe-7851R Field Programmable Gate Array (FPGA) (National Instruments) for hardware control of the system. The system is operated using custom-coded software written in LabVIEW. Current-clamp technique has been used at 1nA with signal filter at 1 kHz. The signal was further digitized by an Axon Instruments Digidata 1322A. The data were then recorded by a LabVIEW 9.0 home-made software, as described previously [9]. While in the medium, a fixed potential of 500 mV was applied to the glucose nanosensor. This biased potential generates an ion current through the liquid-liquid interface that is used as the input into a feedback loop analyzed by the scanning ionic conductance microscope. When a 5–10% current decrease is detected, the glucose sensor is penetrated into the target cell up to 0.8 µm at 100mm/s. The sensor is maintained inside the target cell for a pre-defined time of 60 seconds and then withdrawn to the initial position. Each condition was tested on a minimum of 9 cells. Three consecutive readings were recorded for each individual target cell to ensure reproducibility and robustness. In order to investigate underlying

molecular mechanism of these observed changes, a detailed transcriptome study was performed.

2.2.5 RNA isolation from HD cells

RNA extraction and purification were performed using RNeasy Mini kit from Qiagen. Prior to RNA isolation, HD cells were collected with conventional trypsin treatment and counted using Bio-Rad automated cell counter. A minimum of 5×10^5 cells were used for RNA isolation. Samples were stabilized by treatment with RNAlater (Ambion). After RNA isolation, nucleic acid quantity and quality were checked using NanoDrop (Thermo Fisher Scientific).

2.2.6 RNA-seq library preparation, qRT-PCR and sequencing

2.2.6.1 cDNA synthesis

Extracted total RNA was converted to cDNA using the Smart-seq2 protocol and the cDNA was preamplified as described by a previoud study. Briefly, the cDNA product was mixed with KAPA HiFi HotStart ReadyMix (2X, KAPA Biosystems), ISPCR primer (10uM) and nuclease-free water (Gibco) and PCR amplified as follows: 98 °C 3 min, then 12 cycles of (98 °C 15 s, 67°C 20 s, 72 °C 6 min), with a final extension at 72 °C for 5 min. The cDNA was cleaned up using a 1X ratio of AMPure XP beads (Beckman Coulter).

Table 2.1: Sequences of primers used in this study.

Gene	**Forward**	**Reverse**	**Reference**
SORCS1	5'-AGCCAACAGAAATAAACCTTTCC-3'%	5'-TAATGTGGTCTTCTTCTTGATGT -3'	[49]
ACTB	5'-CACCCGCGAGTACAACCTTC-3'	5'-CCCATACCCACCATCACACC-3'	[50]

2.2.6.2 Quantitative real-time PCR

Real time PCR experiments were performed using 500 ng of reverse transcribed RNA and analyzed with SYBR green qPCR master mix (kapa Biosystems) using an Mx3000P instrument (Stratagene). The relative quantification of Sorcs1 expression levels was performed using a previously described method [48]. For these experiments, *ACTB* was used as a reference gene. *SORCS1* and *ACTB* specific primer sequences designed in previous studies were used (Table 1).

2.2.6.3 Library preparation and sequencing

The cDNA was diluted to ∼300 pg/ul and then processed with the Nextera XT DNA library preparation kit according to the manufacturer's protocol (Illumina). These libraries were purified and size-selected with an average library size of 300–600 bp as determined by Bioanalyzer 2100 high-sensitivity DNA assay (Agilent). Functional library concentration was determined with the KAPA Biosystems library quantification kit. The libraries were denatured after quantification and loaded on Illumina HiSeq 2000 for sequencing.

2.2.6.4 Quality control and mapping of sequencing reads

Sequencing adapter sequences were removed from reads using Cutadapt version 1.4.2 [51]. Quality of preprocessed reads was evaluated using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The preprocessed reads were mapped as paired-end reads using the STAR pipeline [52] with default parameters against the UCSC rn5 Rattus norvegicus genome, combined with the human huntingtin gene (http://hgdownload.soe.ucsc.edu/goldenPath/rn5/bigZips/).

2.2.6.5 Gene expression analysis

HTSeq package was used to generate the gene matrix from the output sam file produced by the STAR alignment. A gtf file containing Ensembl identifiers as well as gene names was used for rat and human genes. For the ERCC genes, the gtf file from Thermo Fisher Scientific was downloaded from the company website (www. thermofisher.com). The gene matrix was then subjected to DESeq2 differential expression analysis, as described [53, 54, 55, 56]. Benjamini-Hochberg multiple testing adjustment was used as a method to deal with p-values that are low due to chance, and not because the measurement represents a significant value. False Discovery Rate values used were calculated by DESeq2 package. Hierarchical Clustering of differentially expressed genes was performed using the Pretty Heatmaps function in R (v. 3.2.3). Pearson coefficients were used as correlation coefficients for the Hierarchical Cluster Analysis heatmaps. The Panther classification system (www.pantherdb.org) was used to visualize the pathways in which differential genes were involved.

2.2.7 Statistical analysis

Analysis of Variance (ANOVA) and independent t-test were performed using Prism (GraphPad) software when indicated.

2.3 Results and discussion

The energy metabolism in presymptomatic and symptomatic HD individuals is known to be defective [57, 58]. Reduced glucose uptake in the striatum and cortex of HD patients prior to the onset of clinical symptoms has been previously revealed by positron emission tomography (PET) scanning. Additionally, this hypometa- bolism demonstrated a high correlation with the onset and progression of HD, independent of the extent of CAG expansion [59]. Recently, Besson et al. reported that the increase of glucose transporters was beneficial to HD pathology in a Drosophila melanogaster model [60]. However, due to the destructive nature of previously employed intracellular glucose measurement methods, a direct interrogation of temporal changes at the transcriptomic and metabolic level could not have been studied. In this work, we took a step forward by deploying functionalized nanopipettes as glucose sensors to measure changes in intracellular glucose concentrations over time, and performed RNA sequencing to evaluate concomitant changes in expression of genes related to glucose metabolism.

2.3.1 Glucose nanosensors reveal temporal intracellular glucose changes at HD cells

Direct monitoring of glucose in cells is difficult due to the small size of the biological specimen and the complexity of the intracellular environment. Glucose nanosensors are made of glucose oxidase (GOx) functionalized quartz nanopipette with ~ 100 nm pores. Due to their small size, glucose nanosensors can be inserted into individual cells, providing a non-destructive and direct *in vitro* quantification tool. We have recently demonstrated the use of this nanosensor in conjunction with a customized cell finder at normal and cancer cells [19] [16]. Typical physiological concentrations of glucose in biological systems are in the low-micromolar range [61]; intracellular glucose levels in HD cells are unknown. More importantly, to our knowledge, there are no reports on the assessment of temporal changes in intracellular glucose concentration in HD models, an assessment that could illuminate one aspect of the cellular pathophysiology of HD. To achieve the goal of measuring intracellular glucose over time, we placed nanosensors in the cytoplasm of wild type (WT) (ST14A-Q15) and HD (ST14A- Q120) cells. Cells underwent consecutive passages in fresh culture medium and glucose measurements were made at the end (day 3 (T1), day 6 (T2), day 11 (T3), day 16 (T4)) of each passage. In the initial experimental conditions, WT and HD cells were cultured in low glucose media (LGM) and high glucose media (HGM) containing 0.5X and 1X (4.5 g/L) glucose respectively. Cells were also cultured in glucose-free media but neither WT nor HD cells survived in this condition. Therefore, measurements could not be conducted

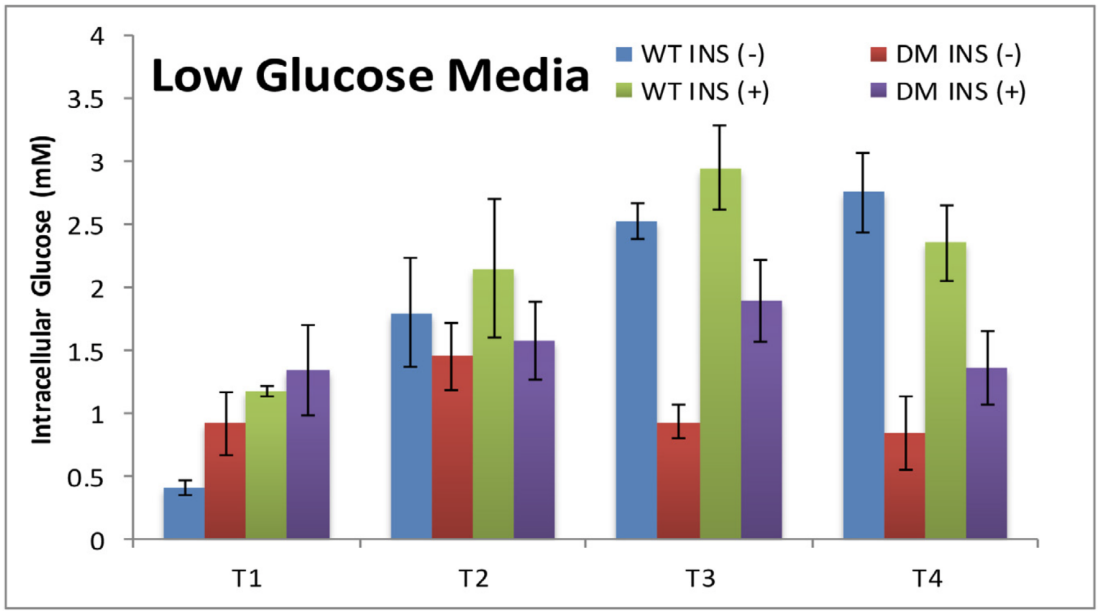

Figure 2.1: Bar graphs showing the average intracellular glucose concentrations measured using GOx-functionalized nanosensor in DMEM with low (LGM) glucose content in the absence of insulin (INS (-)) and 1 nM insulin (INS (+)) in wild type (WT) and disease cell lines (DM). Nine to 12 individual cells with 3 replicates were tested for each condition. The measurements were performed at day 3 (T1), day 6 (T2), day 11(T3), day 16 (T4).

in the absence of glucose. It is important to note that the regular culture condition of these cells does not require or include insulin. Exposure of WT cells to low and high glucose media in the absence of insulin resulted in similar intracellular glucose levels as measured by glucose nanosensors (Figs. 1 and 2).

Intracellular glucose levels were found to be almost stable from T2 to T4, and the range was between 1.79 to and 2.58 mM. The most significant difference was observed for the T1 time point where intracellular levels were about 3 fold lower than those of T2-T4 in wild type cells. This low concentration can be due to either fast metabolism or rapid growth rate of young cells. When the glucose nanosensor was used

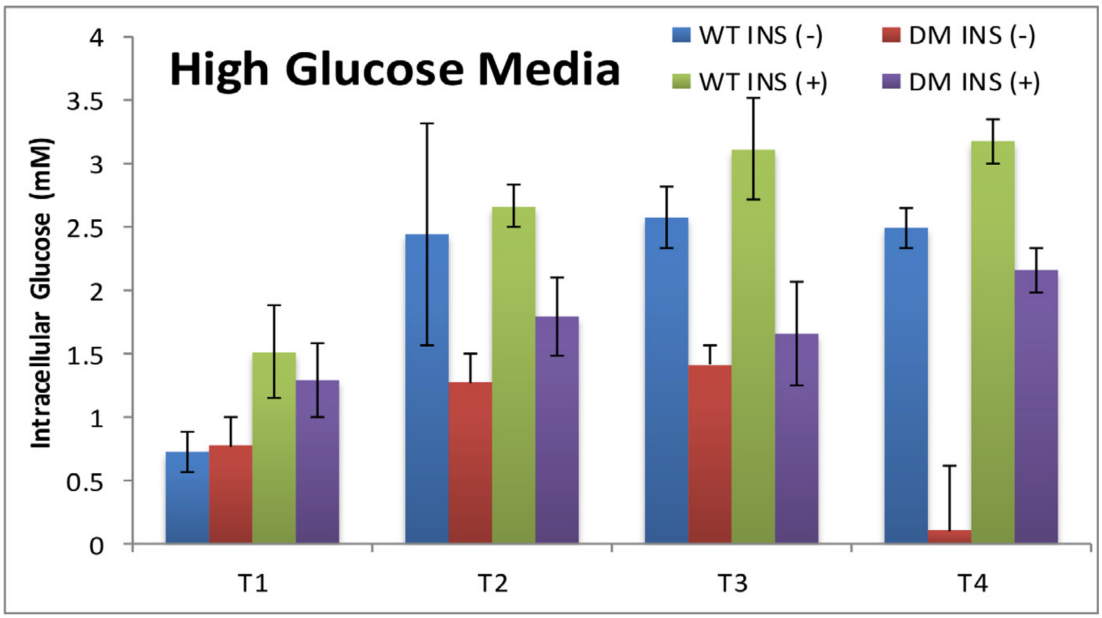

Figure 2.2: Bar graphs displaying the average intracellular glucose concentrations measured using GOx-functionalized nanosensor in DMEM with high glucose (HGM) content in the absence insulin (INS(negative)) and 1 nM insulin (INS (+)) for wild type (WT) and disease cell lines (DM). Nine to 12 individual cells with 3 replicates were tested for each condition. The measurements were performed at day 3 (T1), day 6 (T2), day 11 (T3), day 16 (T4).

on HD cells, we observed that intracellular levels of glucose were about 2.5 fold lower than that found WT cells for T3 and T4 (Figs. 1 and 2).

To compare glucose uptake in WT and HD cells, we exposed the cells to 1 nM insulin (normal non-fasting insulin level), as insulin regulates glucose uptake by activating glucose transporters 1 and 4 (GLUT1 and GLUT4) [62]. We saw a general increase in intracellular glucose levels for all time points for both cell lines (Figs. 1 and 2). As expected, in HGM, intracellular glucose levels were much higher than in LGM. Notably, the effect of insulin was more pronounced at T3 and T4 suggesting that glucose uptake of HD cells is facilitated when insulin is present. The average intracellular

glucose concentrations and their variations for both cell types under all conditions are summarized in Table 2.

2.3.2 Assessing transcriptomics in rat HD cells

We integrated transcriptomics and metabolomics i.e., the metabolome provided phenotypic (glucose) measurements to which we anchored the global measurements of the transcriptome related to glucose pathways. In order to achieve this, RNA-seq transcriptomics was performed to complement the data obtained from glucose nanosensor measurements, using a similar set-up. RNA was extracted from wild type (WT) (ST14A-Q15) and disease (HD) (ST14A-Q120) cells. The gene expression levels were assessed across four passages T1-T4. First, to assess the sequencing by synthesis (SBS) method as a quantitative profiling tool, ERCC spike-in controls were included in the cDNA mix with the samples. ERCC spike-in controls showed Pearson correlation above 90% in all library-to-library comparisons. Also, a 90% Pearson correlation was observed in the expression of ERCC pseudo-genes to the ERCC starting concentrations provided by the ERCC manufacturer (Fig. 3 (A)). Although Spearman correlation was slightly lower (minimum 88%) (Fig. 3 (B)), Pearson correlation is a better parameter for our dataset because it assumes a linear relationship between ERCC gene expression measurements and ERCC starting concentration [63]. Comparisons of upregulated, downregulated and significantly different genes were made with the top 50 genes in a universe of 3213 differentially expressed genes. A recent RNAseq study in human post-mortem HD patients also showed 5480 genes differentially expressed,

Table 2.2: Summary of intracellular glucose measurements at wild type (WT) and disease (HD) cells in the absence I(-) and presence I(+) of insulin. The standard deviations (σ) were calculated for n = 9 replicate measurements for each condition ($\sigma = \sqrt{\frac{1}{N}\Sigma(x_i - \mu)^2}$.

		WT	HD	WT	HD	WT	HD	WT	HD
		T1	T1	T2	T2	T3	T3	T4	T4
[Glucose]	LG I(-)	0.41 (±0.06)	0.92 (±0.25)	1.79 (±0.43)	1.45 (±0.26)	2.53 (±0.13)	0.93 (±0.13)	2.54 (±0.29)	0.84 (±0.29)
[Glucose]	LG I(+)	1.17 (±0.05)	1.34 (±0.35)	2.15 (±0.55)	1.58 (±0.31)	2.61 (±0.16)	1.89 (±0.33)	2.35 (±0.29)	1.36 (±0.29)
[Glucose]	HG I(-)	1 0.72 (±0.16)	0.77 (±0.23)	2.44 (±0.87)	1.28 (±0.23)	2.58 (±0.23)	1.42 (±0.16)	2.49 (±0.23)	0.11 (±0.51)
[Glucose]	HG I(+)	1.52 (±0.36)	1.29 (±0.29)	2.67 (±0.16)	1.80 (±0.31)	3.12 (±0.40)	1.66 (±0.40)	3.19 (±0.18)	2.16 (±0.17)

indicating a similar order of magnitude in the number of differentially expressed genes in rat and humans due to mHTT [37].

2.3.3 Glucose-related genes were differentially expressed in mHTT cells

Specific genes used for the analysis were involved in glycolysis, trichloroacetic acid cycle (TCA), pentose phosphate pathway (PPP) and glucose transport; accessed from a list of genes in glucose metabolism provided in Qiagen's website (www.qiagen.com). The pattern of expression of genes involved in these pathways in mutant and wildtype were plotted as a heatmap using the Pheatmap package available in R [64]. Seven genes in the glycolysis pathway, eleven genes in the TCA, two genes in the PPP, and one gene expressing a glucose transporter were found to be affected by mHTT in this study (Fig. 4).

2.3.4 Survival effect of HK-II and Glut1 on HD cells

Hexokinase catalyzes the first step of glycolysis and represents a regulatory enzyme [65]. Expression of hexokinase II (HK-II) was found to be upregulated in our rat striatal HD cells, a finding also observed by other authors [66]. An increased expression of hexokinase II has also been associated with other neurodegenerative disorders [67] (Fig. 4). It was demonstrated previously that alteration in HK-II gene expression has a direct impact on glucose metabolism [68]. Hexokinases (HKs) phosphorylate the glucose transported through glucose transporters (GLUTs) on the plasma membrane

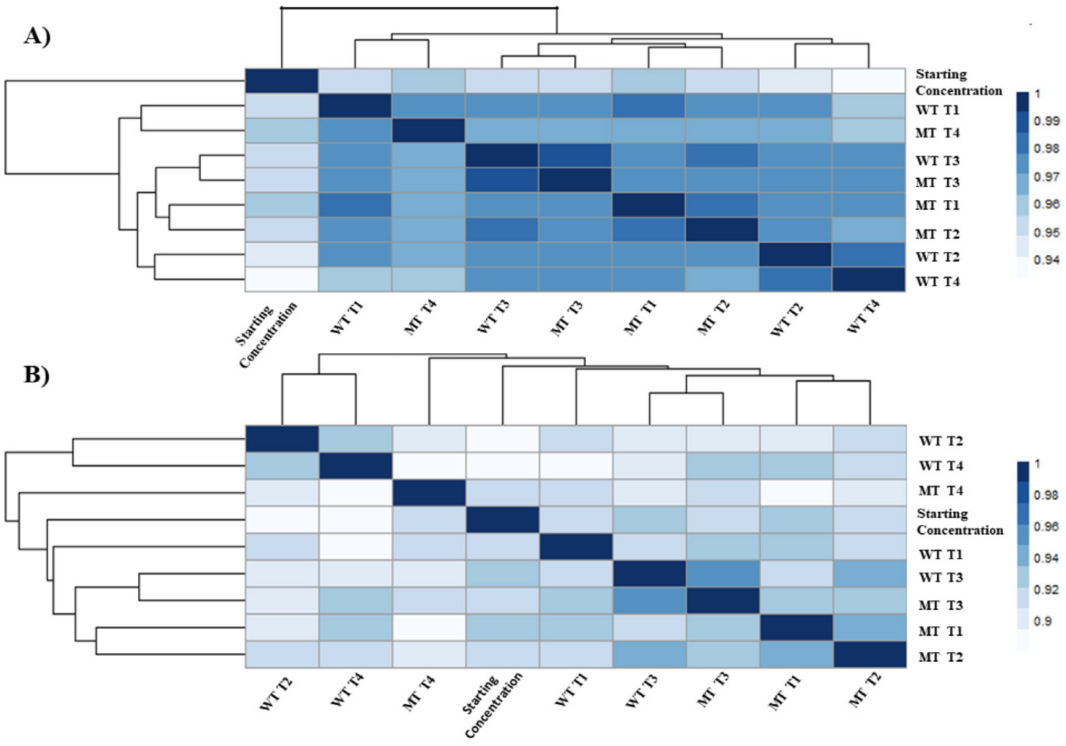

Figure 2.3: A) Pearson correlation of the starting ERCC spike-in concentration and the ERCC spike-in count after library preparation and RNA-seq. ERCC expression values were calculated as log2 of counts. B) Spearman correlation of the starting ERCC spike-in concentration and the ERCC spike-in counts after RNA-seq. ERCC expression values were calculated as log2 of counts.

to produce glucose-6-phosphate (G-6P) (Berg et al., 2002). HK activity is inhibited by G-6P providing a feedback mechanism [69]. The reduced glucose content measured in mutant cells could be attributed to the increased expression of HK-II in HD cells. HK-II upregulation has been previously shown to be an important consequence of metabolic reprogramming in cancer [68]. The expression of Glut1 (Slc2a1), the main class of glucose receptor involved in glucose transport from the blood-brain barrier to the CNS [43], was also upregulated in the mutant cell type (Fig. 4). A previous study in a hematopoietic cell line showed that increased glucose phosphorylation due to co-expression of HK-II and Glut1 has an anti-apoptotic effect. The protective effect was attributed to increase in NADPH activity through the pentose phosphate pathway eliciting an anti-apoptotic effect [70]. The mutant huntingtin protein causes neuronal dysfunction and eventual cell death. Some of the pathways found to be abnormal in Huntington's disease models are transcriptional impairment, neuronal excitotoxicity, oxidative damage, inflammation, apoptosis, and mitochondrial dysfunction. The anti-apoptic effect could slow down the disease progression in our HD rat cell model, similar to observations by the study of Johri and Beal [71]. In summary, our data suggest a compensation mechanism by which increased expression of HK-II and Glut1, may drive mutant cells to survive huntingtin mutation. Insulin treatment is known to increase HK-II mRNA and protein in various cell types and the increase is blocked by inhibition of PI3K, an upstream kinase of Akt, as well as inhibition of mTORC1 [68]. mTORC1 is a protein that integrates signals to regulate cell growth and metabolism. Genes coding for IDH1, ENO3, MDH2 and ALDOC proteins were also found to be differentially expressed (Fig. 4) in mHTT cells.

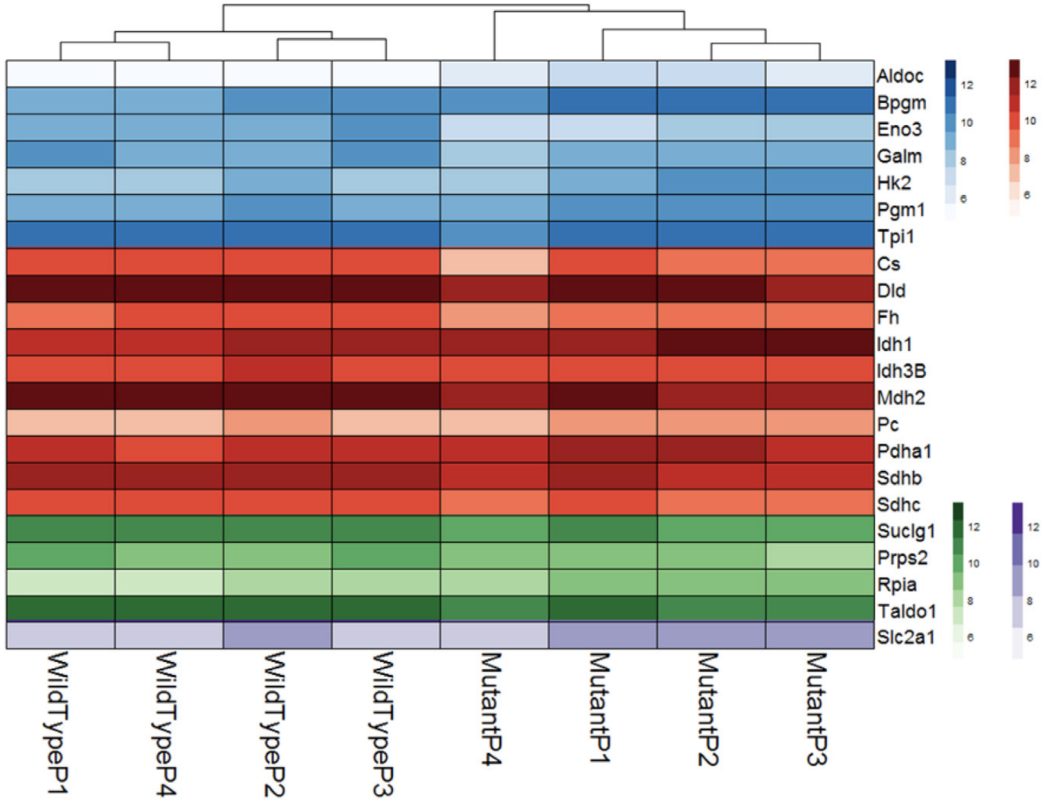

Figure 2.4: Glucose metabolism related genes derived from glycolysis (blue), PPP (red), TCA (green) and glucose transport pathways (purple) differentially expressed in wild type and HD cells. Statistical significance was determined by the Benjamini-Hochberg multiple testing adjustment, as described in DESeq2 documentation, based on p-value comparison. FDR was applied on p-values for significance cut-off.

Eno3 is a gene that codes the enolase version of the muscle. The deficiency of this protein correlates with glycogen storage disease XIII [72]. Together, the results of this glucose homeostasis targeted analysis suggest that glucose metabolism is altered in our HD model, in agreement with other studies [73].

2.3.5 Upregulated and downregulated genes in HD rat cell model

To further understand the perturbation initiated by huntingtin, we analyzed the most abundantly differentiated genes of our cell model in the Panther Classification System (www.pantherdb.org). The group of the top 50 upregulated genes was segmented after the pathway enrichment analysis of genes present using Panther Classification. Fig. 5 shows the heatmap and pie-chart analysis of the up-regulated genes. Fig. 5b shows the main cellular pathways identified by the Panther Classification System among the top 50 expressed genes. They were identified to be cellular receptors controlling signal transduction, including EGF receptor, Wnt signaling, and glutamate receptor. Pathways related to Huntington's and Alzheimer's diseases were also found in the Panther Classification. Igf1 is involved in two pathways: the MAPK pathway, related to cell development and differentiation, and the AKT/PKB pathway that controls insulin metabolic actions (Fig. 5) [74].

SORCS1 protein is necessary for insulin granule secretion in β-cells [20], however was not found annotated with such function by the Panther Classification System in our analysis. SORCS1 is sortilin-related vacuolar protein sorting 10 (VPS-10) domain containing receptor 1. Interestingly, the gene expressing SORCS1 controls protein trafficking and transport of other cellular components, such as the amyloid beta plaques [75, 76] and has been established as an important hallmark of Alzheimer's disease. Gene Sorcs1 was found upregulated in the mutant type in this study (Fig. 5a). A recent study suggested that SORCS1 has a novel regulatory mechanism and acts as a modulator of

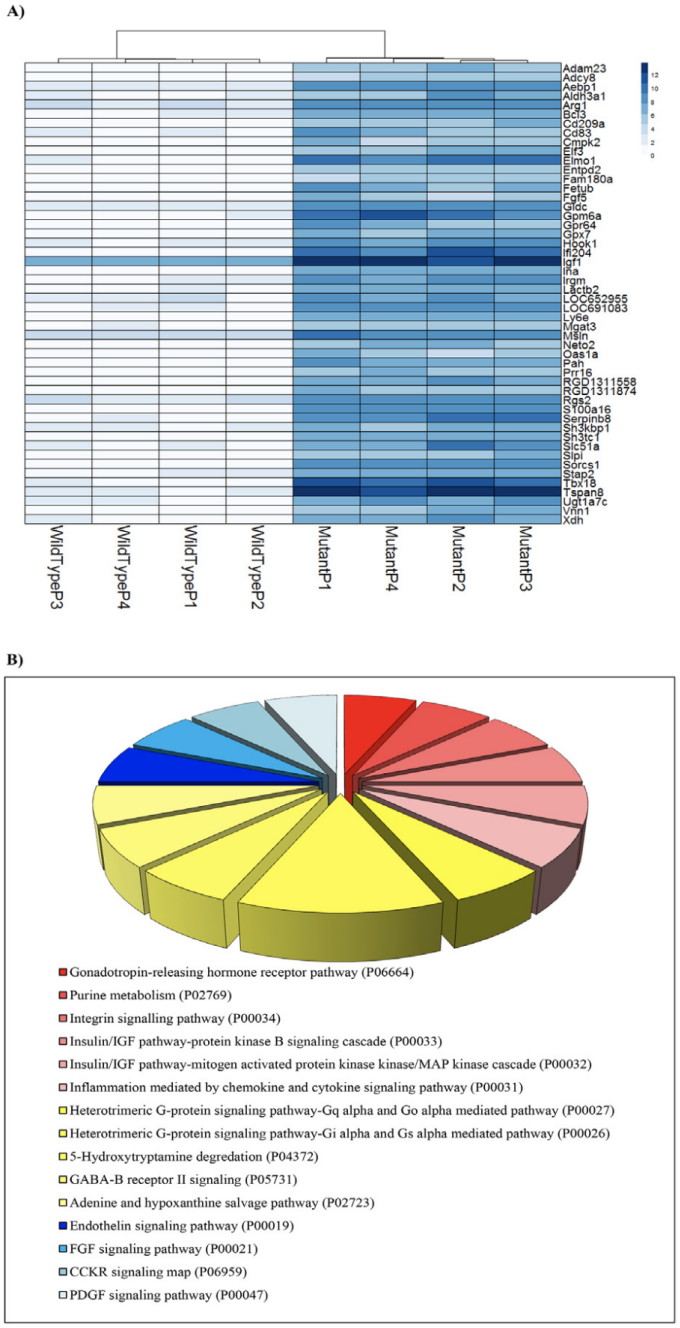

Figure 2.5: (a) Top 50 up-regulated genes in the comparison log2 (value in mutant)/(value in wild-type), as described in the DESeq2 documentation. (b) Pie chart showing Panther pathway analysis of the top 50 up-regulated genes of rat cells expressing human huntingtin.

sortillin function [77]. Sortillin, also a member of the Vps 10 protein sorting receptor family, is involved in biological processes such as glucose and lipid metabolism. It has been proposed that dysfunction of SORCS1 protein may contribute to both the APP/Aβ disturbance underlying Alzheimer's disease (AD) and the insulin/glucose disturbance in Diabetes Mellitus (DM) [78]. To better investigate changes observed in mHTT model in this study, we compared transcriptomics data from other Huntington mice models and included some of the changes observed in AD. Out of 70 genes annotated in the Alzheimer's disease presinilin pathway, 26 genes were found altered in mHTT cells (Figures 1.5 and 1.6).

The presinilin pathway is dysregulated in AD, cleaving several single-transmembrane proteins within the membrane domain, including the amyloid precursor protein, Notch, ErbB4, E-cadherin, Nectin-1alpha, and CD44 (www.pantherdb.org). Notably, another important AD marker, gene Apoe, was affected in our cell model. Interestingly, Vldlr, an Apoe receptor was the top downregulated gene in our study (Fig. 1.7).

As mentioned before, HK-II, a glycolysis enzyme, was affected by the mHTT protein. Besides the already established low-glucose metabolism characteristic of AD, it was observed that AD condition affects the regulation of 6-phosphofructo-2-kinase in human patients, further illustrating common features between AD and HD [79]. Other pathways significantly affected in our mHTT model, including calcium signaling, glutamate receptor activity and synaptic transmission, were observed by other researchers in HD mice and human patients [80, 81, 82, 83]. For instance, Achour and collaborators identified Gata2 transcription factor, a gene involved in establishment of tissue-specific

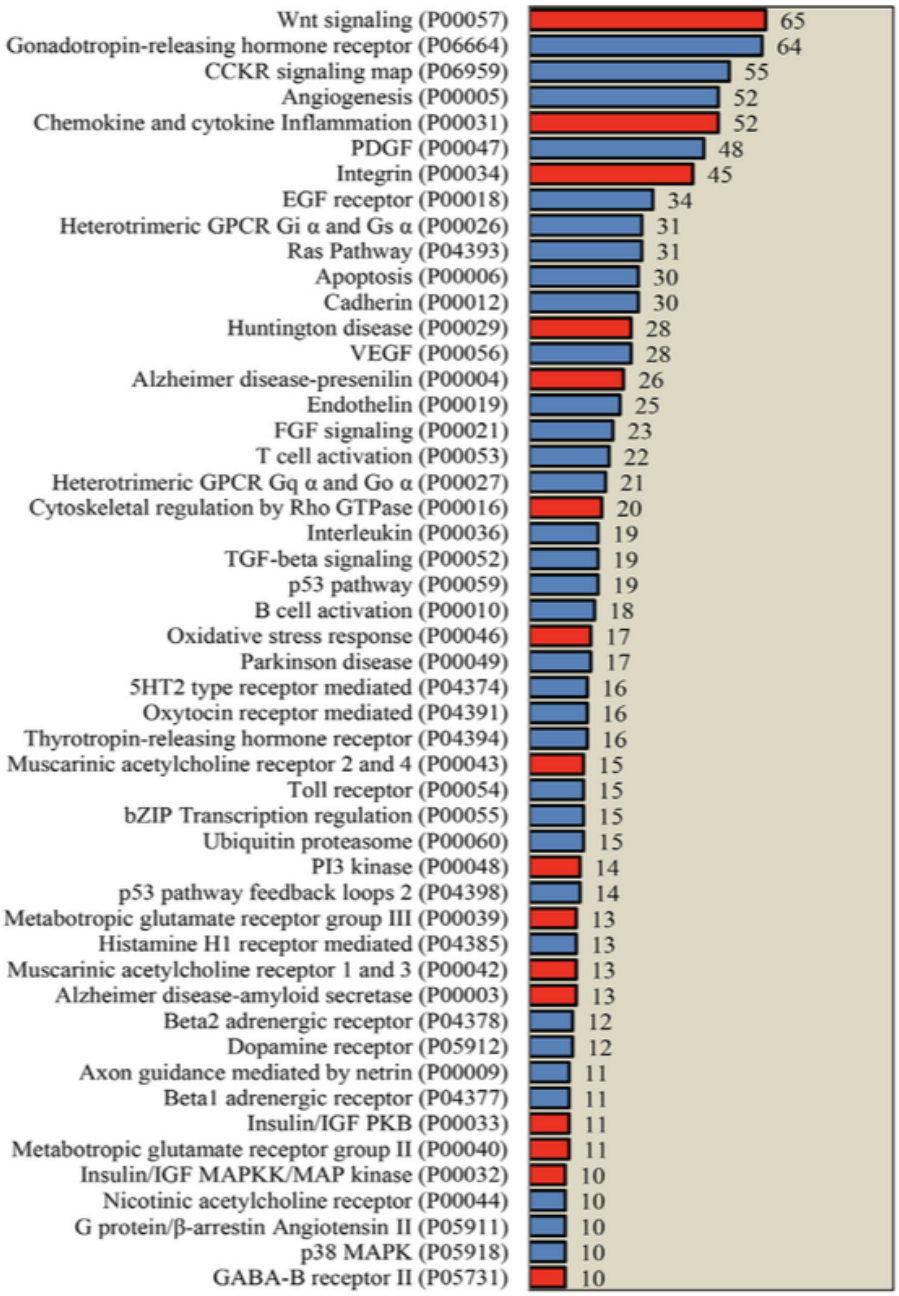

Figure 2.6: Number of differentially expressed genes per signaling pathway according to Pantherdb analysis. A minimum of 10 genes per signaling pathway was requested. Pathways were organized from highest to lowest number of differentially expressed genes within each pathway. Some of the pathways in red bars represent signaling pathways which are discussed in this manuscript.

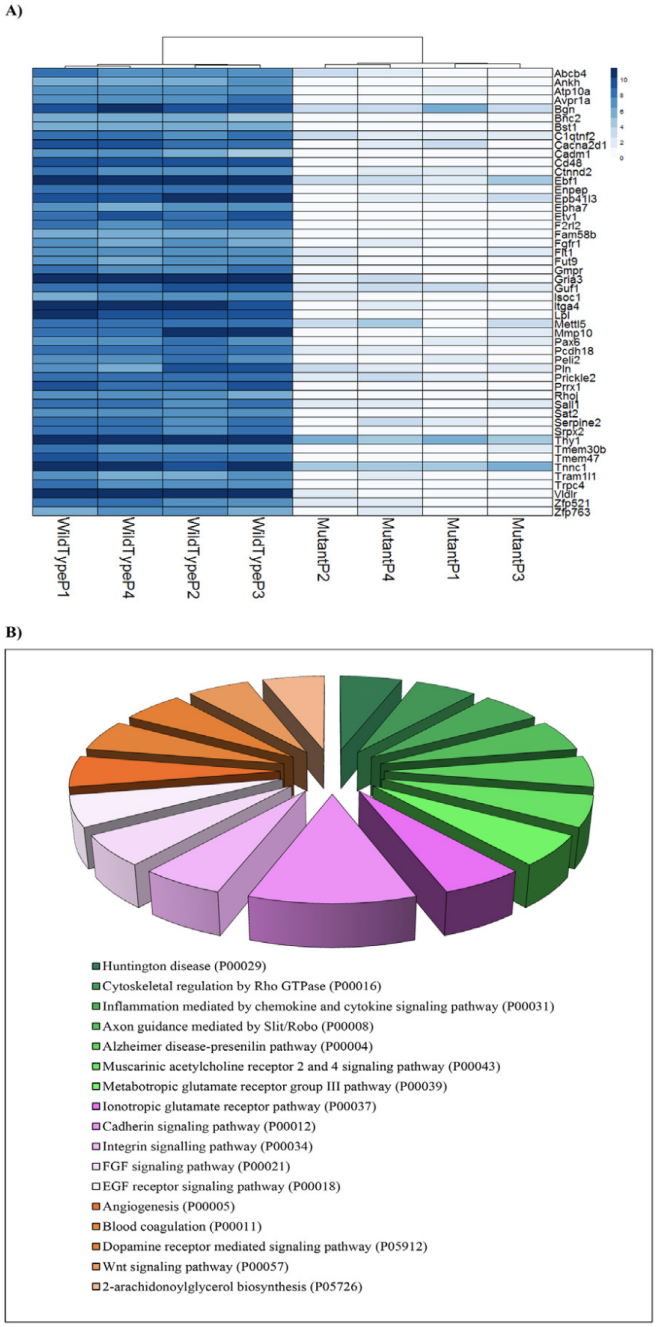

Figure 2.7: Number of differentially expressed genes per signaling pathway according to Pantherdb analysis. A minmum of 10 genes per signaling pathway was requested. Pathways were organized from highest to lowest number of differentially expressed genes within each pathway. Some of the pathways in red bars represent signaling pathways which are discussed in this manuscript.

enhancers, which was found to be affected in the present study, highly expressed in the striatum, and not cerebellum, of HD mice model [84]. Furthermore, genetic variations in intron 1 of *SORCS1* have been associated with Alzheimer's disease in several studies [85, 15]. In order to confirm upregulation of *SORCS1* observed in RNA seq analysis, we stimulated cells with insulin as described in intracellular glucose measurements and performed gene specific PCR analysis (Fig. 8).

SORCS1 was found to be upregulated in mHTT compared to wild type as observed in RNA-seq data. To understand the effect of insulin and/or glucose on expression of *SORCS1* in mHTT cells, qRT-PCR was performed. The mHTT cells not treated with glucose and insulin showed the highest *SORCS1* expression (Fig. 9).

Significant difference was observed when these mHTT cells (no insulin, no glucose) were compared to WT cells grown in all conditions tested. However, we observed no statistically significant difference in *SORCS1* expression in mutant cells treated with either insulin or glucose. Our results suggest that mHTT mutation was the only factor causing higher *SORCS1* expression. However, more stringent and detailed analysis will enable better understanding of the effect of additional glucose and insulin on expression of *SORCS1*. In response to insulin, sortillins can transport GLUT4 to the membrane for uptake of glucose [59] [6]. There are no previous studies demonstrating that the two proteins, SORCS1 and sortilin, can act identically. However, recent literature has indicated that the SORCS1 are sortilin-related CNS expressed proteins, and, as they belong to the same subgroup of VSP10 receptor family, these proteins are related [20, 75, 86]. Since SORCS1 and sortilins belong to the same family of proteins, we were interested

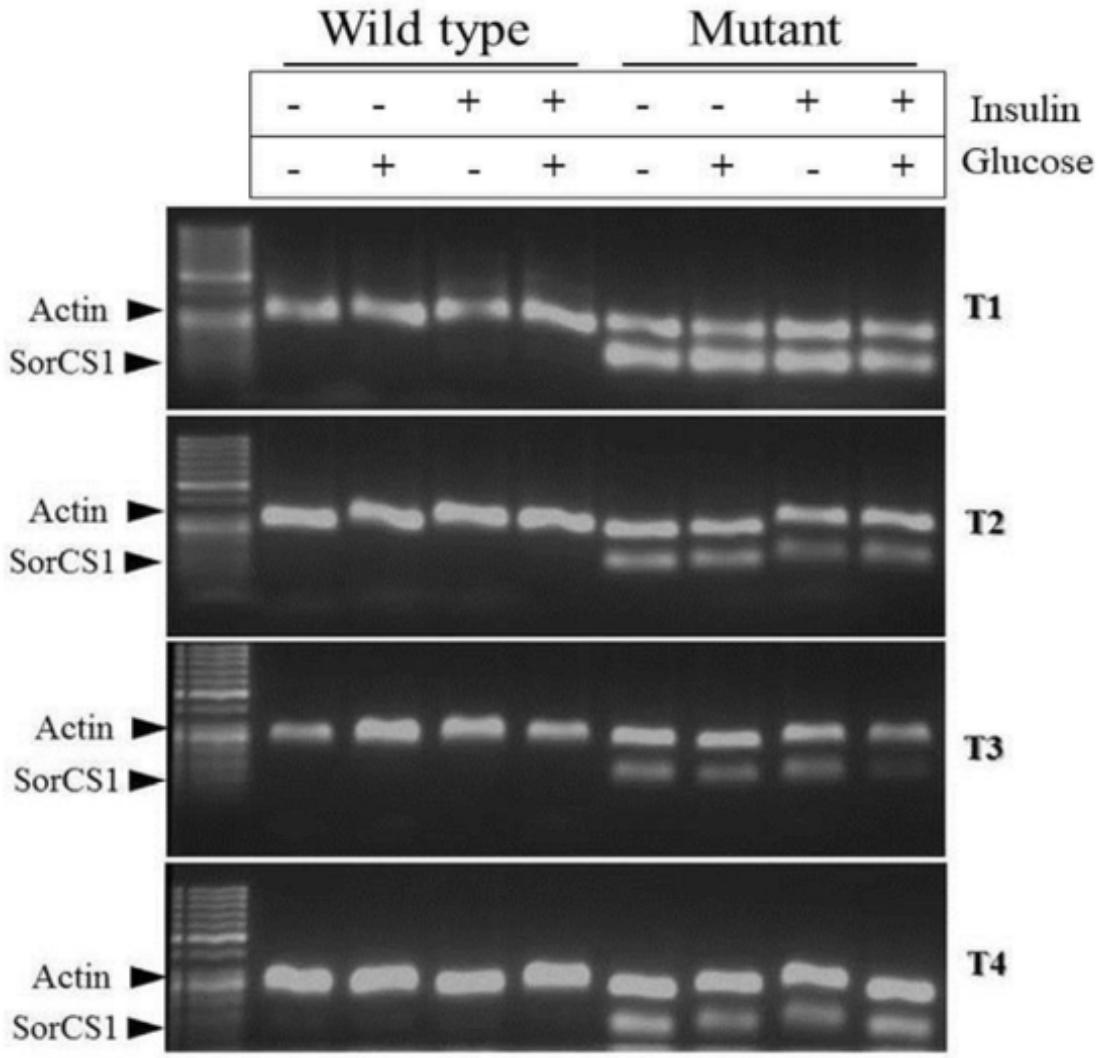

Figure 2.8: Polymerase Chain Reaction products of *SORCS1* and actin primers amplification. 5µL of actin amplicons and 5 µL of *SORCS1* amplicons were mixed and allowed to resolve in agarose gel electrophoresis as indicated above. Insulin and glucose treatments were indicated above each lane. Results suggest a higher expression of *SORCS1* in the mHTT cell type, confirming RNA-seq data.

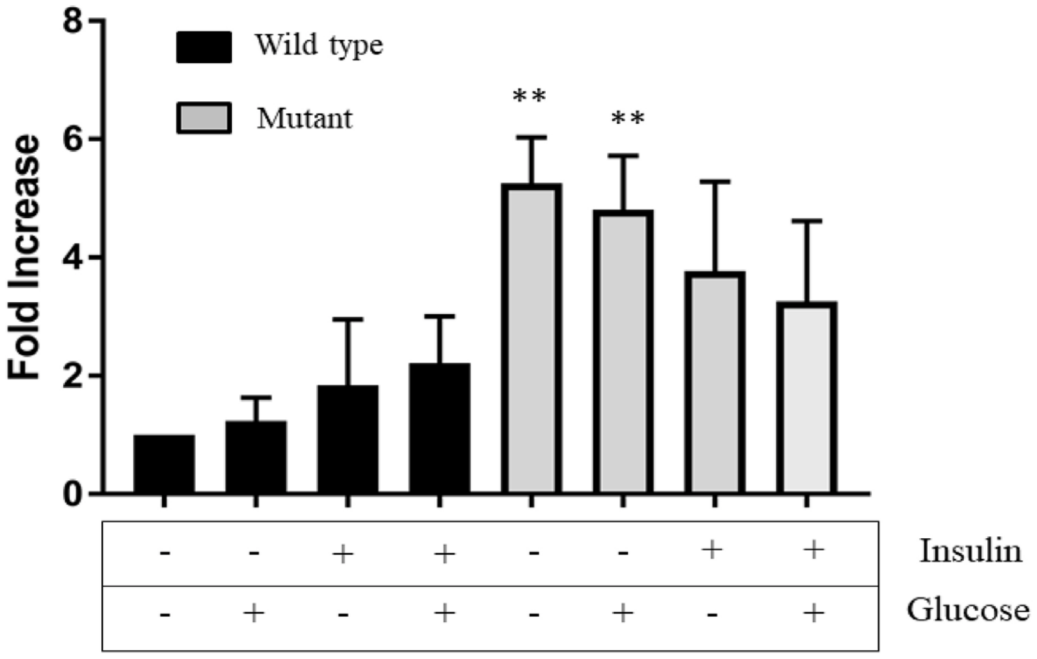

Figure 2.9: Relative gene expression (fold change) of *SORCS1* gene in mHTT and wild type was normalized to reference condition (wild type cells grown in media containing no additional glucose or insulin). *ACTB* was used as endogenous control. mHTT cells had 2- 4-fold increase in *SORCS1* levels. Results are presented as mean (T1, T2, T3 and T4) ± SE. Statistical analysis was assessed by unpaired t-test and one-way ANOVA. **$p < 0.05$ was considered statistically significant.

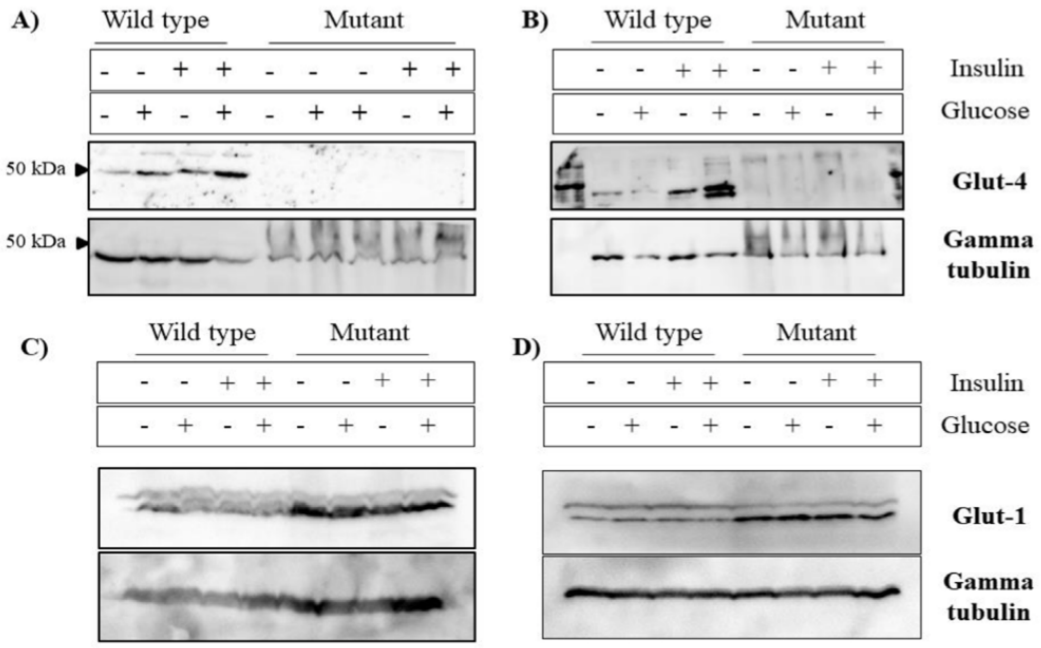

Figure 2.10: Western blot of glucose receptors. Glucose transporter 4 protein expression comparing mHTT and wild type HTT-expressing cell lines corresponding to passages 1 (T1) (A) and 4 (T4), (B) of ST14A cells. A decreased GLUT4 expression is observed due to mHTT. C) Glucose transporter 1 protein expression comparing mHTT and wild type HTT-expressing cell lines corresponding to passages 1 (T1) (C) and 2 (T2), (D) of ST14A cells. Mutants (mHTT cells) had an increased expression of GLUT1 receptor.

in understanding whether increased SORCS1 in cells with Huntington phenotype would affect Glut4 gene expression, thereby increasing glucose uptake. As Glut4 expression was not significantly different in HD and normal cells in the RNA-Seq analysis, we performed protein expression studies. Our results showed that GLUT4 protein was expressed at lower level in rat HD cells (Fig. 10A and B).

Basal levels of GLUT4 along with other glucose receptors including GLUT1, the main class of glucose receptor, may compensate for the lack of GLUT4 protein observed in mHTT cells. Nevertheless, other factors including post-transcriptional modifi-

cation or action of miRNA may also have an effect on decreased GLUT4 protein levels, which needs to be investigated further. Hou and Pessin reported that overexpression of sortillin stabilizes GLUT4 protein, increases the formation of insulin responsive compartments, and promotes insulin-stimulated glucose uptake [6]. Higher GLUT4 and/or higher metabolism could explain the increased intracellular glucose concentration in wild-type cells and a time-dependent study could give a better idea whether GLUT4 is indeed regulated by SORCS1 in addition to sortillin. The role of sortilins in the translocation of GLUT4 proteins was revised in the study of Hou and Pessin [6]. As GLUT4 levels were lower in mHTT cells, we became interested in analyzing GLUT1 levels, the main class of glucose receptor. Results showed higher GLUT1 levels in mHTT cells suggesting that GLUT1 could indeed compensate for the lack of GLUT4 in mHTT cells (Fig. 10C). The fine regulation of brain developed neutrophic factor (BDNF) by sortillin has been implicated in neuronal and tumor cell survival. Interestingly, BDNF protein was also shown as a diabetes and Huntington's disease marker [30, 87]. In our cell model, following the trends described by Zuccato and collaborators, BDNF levels were lower in mutant cells than in wild type. Furthermore, it has been demonstrated that BDNF protein is one of the molecules that is a target of organelle sorting mediated by SORCS1 [88]. BDNF, once transported to the striatum from the cortex, induces cholesterol synthesis and is related to synaptic plasticity and neuronal survival [89]. Sortillin has also been implicated in LDL- cholesterol metabolism and VLDL secretion [90, 91]. The very low density lipoprotein receptor (VLDLR) was found as the top downregulated gene in HD cells as compared to wild type cells in our data set (Fig. 7). VLDLR

is an apolipoprotein E (APOE) receptor and the interaction of these two proteins has important implications in lipid metabolism [92]. Downregulation of *VLDLR* gene indicates that lipid metabolism is also impaired in mutant huntingtin striatal cells, as has been shown by another group [93]. Other genes that were found to be downregulated in this study are shown in Fig. 7a, along with the pathways they belonged to in Fig. 7b. Together, these data suggest the involvement of *SORCS1*, insulin, *GLUT4* and *GLUT1* in the HD condition, at least in our cell model, providing molecular candidates that support a decreased intracellular glucose concentration due to mutant huntingtin. To our knowledge, no other group has reported the involvement of *SORCS1*, an Alzheimer's disease and diabetes marker, in a Huntington's disease model. *SORCS1* may not only be involved in impairment of glucose metabolism, but also lipid metabolism in Huntington's disease.

2.3.6 Other pathways affected by mutant huntingtin

Using transcriptomics and metagenomics, we sought to understand how mHTT influences not only HD related pathways but also other major pathways. In order to understand the pathways affected, we classified differentially expressed genes according to their signaling pathways, with at least 10 genes observed per pathway in the Pantherdb classification system. 1145 genes (out of 3213) are represented in Fig. 6. The signaling pathway most significantly affected by mutated huntingtin in number of genes was the wingless-related (Wnt) signaling pathway (P00057), showing 65 disregulated genes. Beta-catenins, a canonical element from the Wnt pathway, forms a

complex with GSK-3B; they regulate epidermal growth factor and insulin, as well as control cell fate. Other genes that were differentially expressed were important in the insulin/MAPK/PKB/IGF pathways (P00032 and P00033). 64 genes were impaired in the gonadotrophin-releasing hormone receptor (P06664). Dysregulation of inflammation mediated by the cytokines and chemokines pathway (P00031) supports the anti-inflammatory treatments sug- gested for human HD [37]. Glutamatergic pathways were also found to be affected in our study (P00037 and P00039). Glutamate ionotropic receptor AMPA type subunit 3 (GRIA3), a gene in this pathway, was downregulated in our study (Fig. 7). Morton et al. also showed this gene to be downregulated in R2/6 mice, affecting synaptic plasticity [94]. A recent study showed SORCS1 regulating the AMPA receptors [75]. All the other pathways affected due to mutant huntingtin (containing 10 or more than 10 genes per pathway) are shown in Fig. 6. In summary, data shown on Fig. 6 support general observations from the literature about Huntington's disease, as well as provide support to our observation of SORCS1 being dysregulated due to huntingtin mutation, in that this gene interacts with glutamate receptors [75].

2.4 Conclusion

Correlation between HD and glucose blood levels has been extensively studied. Although some studies do not show a correlation between HD and glucose metabolism, [31], other studies suggest the involvement of insulin resistance mediated by the huntingtin mutation and cross-talk between glucose metabolism and other metabolic path-

ways. Among these pathways, involvement of fatty-acid biosynthesis and cholesterol synthesis are described in HD progression [89, 93, 95, 96, 97, 98, 99]. Here, we showed impaired glucose metabolism in rat striatum cells using a functionalized nanopipette as a non-destructive intra-cellular glucose measurement technique. Impaired glucose metabolism along with altered expression of genes concerned with energy metabolism contributes to HD pathogenesis. A new additional target, SORCS1 was identified to be altered in HD cells along with previously reported transporter related genes, GLUT1 and GLUT4. Increased levels of SORCS1 correlate positively with loss of the ability of the cell to sort proteins like GLUT4 properly. Further work on generating a knock-out mutant of *SORCS1* gene will shed more light on better understanding the role of this gene in Huntington's disease. GLUT1 and HK-II expression may have a beneficial effect on HD pathology by delaying apoptosis. The induction of genes in the PPP could possibly delay HD progression, thereby providing efficient neuroprotection against reactive oxygen species. These data represent a comprehensive integration of metabolic and transcriptomic expression data in rat HD cells. This research yields important advances in our understanding of HD pathogenesis. It remains to be shown whether the changes observed here are similar across different cell types within the brain of affected human individuals. In Chapter 3, we start addressing the involvement of genes and/or proteins identified in Chapter 2 (e.g. sortilins, tetraspanins and MHC) in human Huntington's disease. We specifically highlight the aspects of these genes that are involved in the immunological function.

Chapter 3

Mutant Huntingtin Affects Diabetes and Alzheimer's Markers in Human and Cell Models of Huntington's Disease

Overview

The following is a transcript of the Chaves *et al.*, (2019) paper [27]. I decided to explore the idea that sortilins, among the top hits of the RNA-Seq analysis I performed in ST14A cells, were involved in human HD. To do that, I performed a Genome-Wide Association Study, where I correlated human HD and SNPs in sortilin genes. In Chapter 3, we start addressing the involvement of genes and/or proteins identified in Chapter 2 (e.g. sortilins, tetraspanins and MHC) in human Huntington's disease. I further explore the concept that sortilins have an immunological function, as do MHC and tetraspanins. These ideas are particularly important in explaining type 1 diabetes phenotypes in HD.

Abstract

A higher incidence of diabetes was observed among family members of individuals affected by Huntington's Disease with no follow-up studies investigating the genetic nature of the observation. Using a genome-wide association study (GWAS), RNA sequencing (RNA-Seq) analysis and western blotting of *Rattus norvegicus* and human, we were able to identify that the gene family of sortilin receptors was affected in Huntington's Disease patients. We observed that less than 5% of SNPs were of statistical significance and that sortilins and HLA/MHC gene expression or SNPs were associated with mutant huntingtin (mHTT). These results suggest that ST14A cells derived from *R. norvegicus* are a reliable model of HD, since sortilins were identified through analysis of the transcriptome in these cells. These findings help highlight the genes involved in mechanisms targeted by diabetes drugs, such as glucose transporters as well as proteins controlling insulin release related to mHTT. To the best of our knowledge, this is the first GWAS using RNA-Seq data from both ST14A rat HD cell model and human Huntington's Disease.

3.1 Introduction

3.1.1 Biology of Huntingtin and Identification of DNA Polymorphism Causing HD

Huntington's disease (HD), a neurodegeneration characterized by motor, cognitive and psychiatric symptoms is caused by an unstable expansion (CAG) in a polyg-

lutamine (polyQ) tract in the N-terminal of the huntingtin (HTT) gene. This condition is triggered when the number of CAG repeats on exon 1 of HTT exceeds 39, while 36-39 CAGs repeats result in an uncertain incidence of the disease [10]. Features of the genetic inheritance of CAG repeats in HTT were unknown for more than 200 years; George Huntington described Huntington's chorea in 1872 and linkage mapping of the genomic region to HD occurred in 1993 [10, 12]. HTT is highly conserved from flies to mammals and the N-terminal is the most extensively studied region of the protein due to the CAG expansion on exon 1. HTT also contains HEAT repeats which play an important role in protein-protein interactions. HEAT repeats are a common helical motif in the Huntingtin protein, Elongation Factor 3, protein phosphatase 2A, and TOR1 [100]. Tools such as the restriction fragment length polymorphism (RFLP) have been extensively used in determining DNA polymorphisms of HTT inherited in HD families [11], but rarely used to study other parts of the genome of HD individuals. RFLP was used in linkage disequilibrium studies that characterized the genetic pattern of CAG repeats in HTT in HD families. The reason for the lack of studies on the overall genome of HD individuals after characterization of mutant HTT is the assumption that no other genomic region influences human HD more than HTT. A study of SNPs in different parts of the genome of HD individuals may help explain different genetic phenomena including increased prevalence of diabetes among families affected with HD and cases where the CAG abnormal expansion was not present in the father nor in the mother, but manifested in the descendants [95, 101, 102]. Due to the great number of HTT molecular partners in events such as DNA and RNA metabolism, endocytosis,

subcellular tra cking and cellular homeostasis, genetic diversity of several proteins that interact with HTT may affect prevalence these conditions among HD patients and their families. Saudou and Humbert reviewed the interaction of HTT with proteins of the molecular motor machinery such as dynein, dynactin and kinesin and the role of HEAT repeats allowing HTT to serve as a scaffolding protein in these biological processes [2]. Zala et al., from the same group, found that the HTT-dependent transport of GADPH, a gene from the glycolysis pathway, provides energy to sustain the fast transport of vesicles using the cytoskeleton [103]. These studies support the notion that HTT acts as a protein important in the transport of vesicles, also influencing cellular endocytosis and exocytosis.

3.1.2 Neurodegeneration and Metabolic Diseases

Endocytosis and exocytosis are cellular mechanisms that maintain the homeostasis of a cell and play a direct role in neurodegenerations and metabolic diseases. Endocytosis and exocytosis influence Type 1 Diabetes (T1D) and Type 2 Diabetes (T2D) in different ways. T1D or insulin-dependent diabetes, features autoantibodies attacking proteins in the membrane of pancreatic cells. Problems in the processing of antigens transported intracellularly by components such as phagosome, proteasome, Golgi apparatus, ER, endosomes and the cell membrane are often associated with T1D. It has been observed that glutamate decarboxylase, zinc transporter-8, IA-2, tetraspanin 7 and insulin act as humoral autoantigens in T1D [104]. T2D, on the other hand, is characterized by a lack of response to insulin in cells that need to use glucose, such as

muscle cells, liver cells and adipocytes, even with the production of the hormone in the pancreas. In cells of patients with T2D, glucose transporters and proteins involved in the translocation of glucose transporters do not correctly respond to the insulin signal. This results in a poor cellular uptake of glucose from the blood stream. Among components involved in the translocation of glucose receptors to the cell membrane, sortilins control sorting of proteins across different cellular compartments, including traffic between the Golgi apparatus and the cellular membrane [6]. Translocation of insulin to the secretory pathway following the release of the hormone in the blood stream is a cellular role performed with assistance of sortilins [20]. The first report of mHTT affecting diabetes phenotypes was made in 1972, when a higher prevalence of diabetes was detected among HD patients [102]. This observation was later supported by findings indicating increased chances of diabetes among family members of individuals affected with HD [95]. Since HTT is a scaffolding protein that connects cytoskeleton proteins to proteins of the endocytosis and exocytosis pathways, it is possible that mutant mHTT affects diabetes phenotypes related to these pathways.

3.1.3 VPS10P-Domain Receptors or Sortilins: Regulators of Subcellular Protein Trafficking and Markers of Diabetes and Neurodegeneration

The vacuolar protein sorting 10 protein (VPS10P) domain is a 700-amino-acid motif first identified in *Saccharomyces cerevisiae* that directs the tra cking of lysosomal enzymes from the Golgi apparatus to the vacuole [105]. VPS10P-domain receptors

in mammals contain five members; SORT1 (sortilin), SORLA, SORCS1, SORCS2 and SORCS3. These receptors bind and neutralize a variety of ligands such as trophic factors, neuropeptides, glucose receptor 4 (GLUT4) and the amyloid beta peptide between the trans-Golgi network, the endosomes and the plasma-membrane [6, 14, 8, 15]. In macrophages, SORT1 binds other proteins from the Golgi apparatus causing the migration of vesicles using the retromer complex. Vesicles containing sortilins end up fusing to the phagosome of macrophages and are known to deliver specific proteins required for immunological control of *Mycobacterium tuberculosis* by macrophages [21]. Due to its involvement with subcellular transport and signaling transduction, Reuter referred to SORT1 as a dual function protein. As a transduction protein, SORT1 controls the translocation of epidermal growth factor ligands. As an intracellular protein transport controller, SORT1 acts in antigen presentation in dendritic cells, illustrating the role of sortilin genes with the immune system [106]. Considering the involvement of sortilin genes in immunity mediated by macrophages, sorting of cellular proteins, as well as our previous study indicating the up-regulation of sortilin SORCS1 in a cell model expressing mHTT [6, 20, 21, 28, 106], here focused on investigating the association of sortilins with Huntington's Disease using human and rat RNA-Seq datasets. We explore the involvement of sortilins in an antigen presentation in HD and specifically provide evidence that proves sortilins are part of the genetic component leading to diabetes in human HD.

3.2 Materials and Methods

3.2.1 Library Preparation and Sequencing

ST14A total RNA was extracted and converted to cDNA using the Smart-seq2 protocol [107]. cDNA was then processed with the Nextera XT DNA library preparation kit according to the manufacturer's protocol (Illumina, San Diego, CA, USA). Libraries were purified and size-selected using the Nvigen Size Selection Kit according to the manufacturer's protocol [108]. The bioanalyzer 2100 high-sensitivity DNA assay (Agilent, Santa Clara, CA, USA) was used to check on the size range. The functional library concentration was determined with the KAPA Biosystems library quantification kit. The libraries were denatured after quantification and loaded on Illumina HiSeq 2000 for sequencing.

3.2.2 Human and Rat Genomes

We used the human genome assembly version GRCh37.p13 (hg19) available at the Ensembl Genome Browser to align the human sequencing files analyzed in this study. The Rattus norvegicus reference genome version Rnor5.0 was used to map rat sequencing files. Gene coordinates for isolation of genetic variants using Unix commands and dbSNP IDs were also acquired using the Ensembl Genome Browser website, when indicated.

3.2.3 Genome-Wide Association Study (GWAS)

The genome-wide association of RNA-Seq reads with Huntington's disease was performed using a costumer Unix/bash script disclosed by Mohammed Khalfan of the New York University Genetics Core Facility. Briefly, reads were aligned to the reference genome using BWA, sorted, converted to a BAM file, duplicate marked and indexed using Picard. The genome analyses toolkit (GATK) developed and provided by the Broad Institute of Harvard and MIT was then used for realignment, variant calling, single nucleotide polymorphism (SNP), indel filtering and extraction, base quality score recalibration, and covariate analyses [109, 110, 111]. Knowing that GWAS strategies are relatively new in science here we use the notion of GWAS as revised by Visscher: "GWAS is an experimental design used to detect associations between genetic variants and traits in samples from populations", to determine that the algorithm used by this report represents the first GWAS using RNA-Seq data from the HD cell model and human Huntington's Disease [112]. Outputs of the GATK/GWAS pipeline were files including the variant call format (VCF), containing the SNPs identified for processing described and reported in this document. A list of Unix commands used for analysis shown in this study are provided in the Supplementary Materials.

3.2.4 Gene Expression Omnibus (GEO) RNA-Seq Datasets

We used GEO datasets previously published by Labadorf et al. (2015) (GEO accession number GSE64810) [37], Lin et al. (2016) (GEO accession number GSM2100621) [113] and HD iPSC Consortium (2017) (GEO accession number GSE95344) [114]. In

this manuscript, we refer to these datasets as Labadorf (2015), Lin (2016) and HD iPSC Consortium (2017) datasets respectively (publication year may or may not be indicated). Files .vcf and .count generated respectively by the GATK and DESeq2 pipelines are included in the Supplementary Materials (these files contain respectively the DNA genetic variants and the read counts detected by these pipelines).

3.2.5 Post-GATK/GWAS Processing

Several processing steps were required after completion of the GATK/GWAS pipeline. These steps were necessary to allow formatting of the output files for compliance with downstream software requirements and were accomplished using Unix commands depicted in the Supplementary Materials (this section includes a brief description of each Unix command used as well as extra results supporting findings described in the main document). Other than the Unix commands depicted in the Supplementary Materials, bcftools (version 1.9) [115], vcftools (version 0.1.14) [116] and PLINK [117] were used to adjust the format of files used in the analysis. PLINK was used for the extraction of PED and MAP files.

3.2.6 Statistical Analysis

PLINK was used for statistical testing of the comparison cases versus controls. Fisher's exact test was calculated as depicted in the command shown in the Supplementary Materials to compare the abundance of alleles in the cases and control individuals [117]. A p-value threshold of 0.05 was used to consider a genetic variant statistically

significant in our GATK/GWAS pipeline.

3.2.7 Manhattan Plots

R package Qqman was used to plot p-values below 0.05, indicating the association between HD and the specific genetic variant detected in each genomic position, according to package instructions [118]. We highlighted SNPs located outside of the gene body up to 1,000,000 base-pairs upstream and downstream of the gene body. We considered that within this region there existed regulatory elements influencing the genetic products encoded by the gene body.

3.2.8 Linkage Visualization Using Haploview

PLINK was used to extract pedigree and .map files from .vcf files and to input Haploview using default parameters, as determined by the Broad Institute guidelines in the website and original paper recommendations [119]. The haplotype block was determined using the Four Gametes Rule option available on Haploview. Linkage plots were constructed with the standard LD color scheme (D0/LOD). Color represents linkage, in the following order: red (LOD 2 and D0 = 1), pink (LOD 2 andD0 ¡1),blue(LOD¡2andD0 =1),andwhite(LOD¡2andD0 ¡1).

3.2.9 Validation of GATK/GWAS DNA Variants Identification by Pyrosequencing

The NGS Pyrosequencing method was used as a genotyping validation of the GATK/GWAS algorithm used in this study as previously described [120, 121].

3.3 Results

3.3.1 mHTT Is Associated with SORCS1 Protein Up-Regulation and Sortilins SNPs

To investigate whether genetic variants of sortilin genes were associated with human HD, we performed a GWAS in rat and human datasets using the Genome Analysis Tool Kit (GATK). Table 2.1 represents an overview of the significant SNPs detected in the entire RNA-Seq dataset as well as in the sortilins gene body and vicinity.

Based on the total number of SNPs identified, we were able to distinguish two groups of datasets; those with low genetic variance and datasets with high genetic variance. It was observed that 220,000 Single Nucleotide Polymorphisms were significantly associated with mHTT in the ST14A and HD iPSC Consortium data, whereas more than 2.6 million SNPs were significantly associated with mHTT in the Labadorf and Lin datasets (Table 1). 1.4–3.80% of all SNPs detected were significant in each dataset and this similarity indicates that variants were consistent across all groups that were analyzed. We hypothesized that SNPs near the genes in the study could influence gene expression levels. To substantiate our hypothesis, we counted the number of SNPs in

Table 3.1: Summary of DNA variants detected in GATK/GWAS pipeline of rat and human RNA-Seq. SNP variants locate in gene body and regulatory vicinity.

Dataset	**Total SNPs**	**Significant**	**Sortilin**	**SNPs in Sortilins**
R. norvegicus	222997	8464 (3.80%)	*SORCS1*	13
			SORCS2	8
			SORCS3	2
			SORT1	11
			SORL1	5
Labadorf	2658838	84349 (3.17%)	*SORCS1*	8
			SORCS2	171
			SORCS3	53
			SORT1	153
			SORL1	38
Lin	3172675	44669 (1.41%)	*SORCS1*	9
			SORCS2	121
			SORCS3	23
			SORT1	111
			SORL1	41
HD iPSC Consortium	238013	8979 (3.38%)	*SORCS1*	0
			SORCS2	20
			SORCS3	3
			SORT1	22
			SORL1	6

the gene body and regulatory regions. For this analysis, the gene regulatory region was defined as a body of $\pm 1,000,000bp$ nucleotides away from the gene body. We used a custom bash script to isolate the SNP positions (Supplementary Material, Command 1A). In the ST14A rat dataset, only variants of the *SORCS1* gene were identified across the five possible sortilins (six SNPs in the gene body, 13 SNPs in the regulatory region) (Table 2.1). For a better visualization of sortilins SNPs observed in the GATK/GWAS pipeline, a plot was developed using the positions of the mutations relative to the sortilin gene organization and the dataset in which SNPs were identified as shown in Figure 2.1.

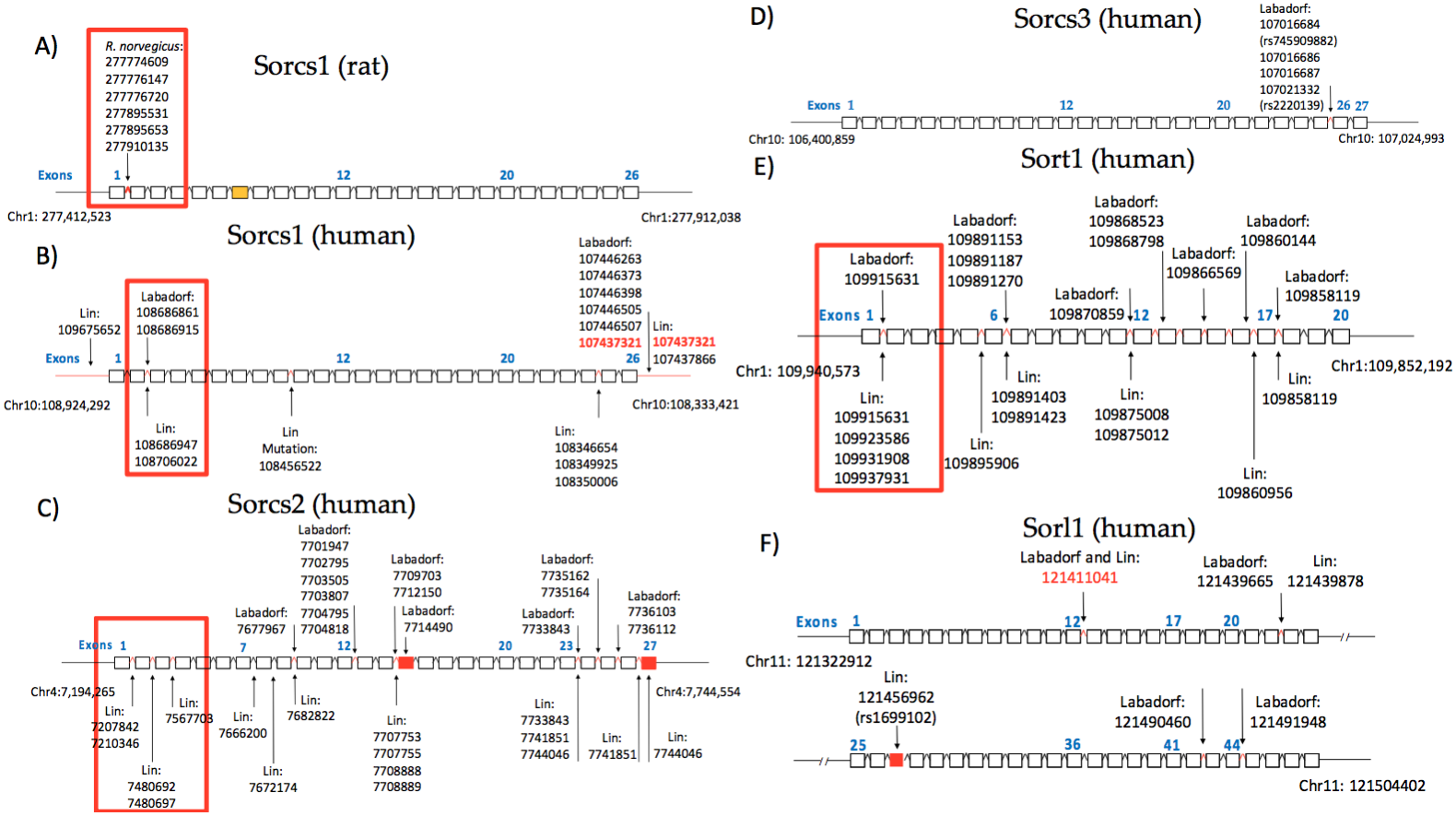

Figure 3.1: Organization genomic regions of sortilins showing SNPs detected by the GWAS/GATK pipeline across introns and exons. The dataset of the identification is indicated. A) *SORCS1* gene organization in *R. norvegicus*. B) *SORCS1* gene organization in human. C) *SORCS2* gene organization in human. D) *SORCS3* gene organization in human. E) *SORT1* gene organization in human. F) *SORL1* gene organization in human. Mutations inside the red squares represent SNPs in introns organization in human. Mutations inside red squares represent SNPs in introns between exons that encode the VPS10P domain [8]. Exons are indicated with their numbers in blue for identification.

From these results, it can be observed that sortilins had mutations significantly associated with Huntington's Disease (p-value < 0.05, Fisher's exact test comparison) in the Labadorf dataset. Only *SORCS3* and Sortilin-related (*SORL1*) did not show mutations flanking introns two and three, which are located between the exons encoding the VPS10P domain (Figure 2.1). For a list of all positions of significant SNPs associated with HD in this study, see the Supplementary Materials PLINK output. Figure 2A represents the agarose gel electrophoresis results. From the gel it is evident that the SORCS1 protein levels were up-regulated in the mHTT-expressing cells, a phenomena that was previously suggested by our group [28]. To further access the effect of mHTT on the biology of sortilins, we investigated SNPs (Figure 2B) and the gene expression (Figure 2C) of sortilins associated with HD.

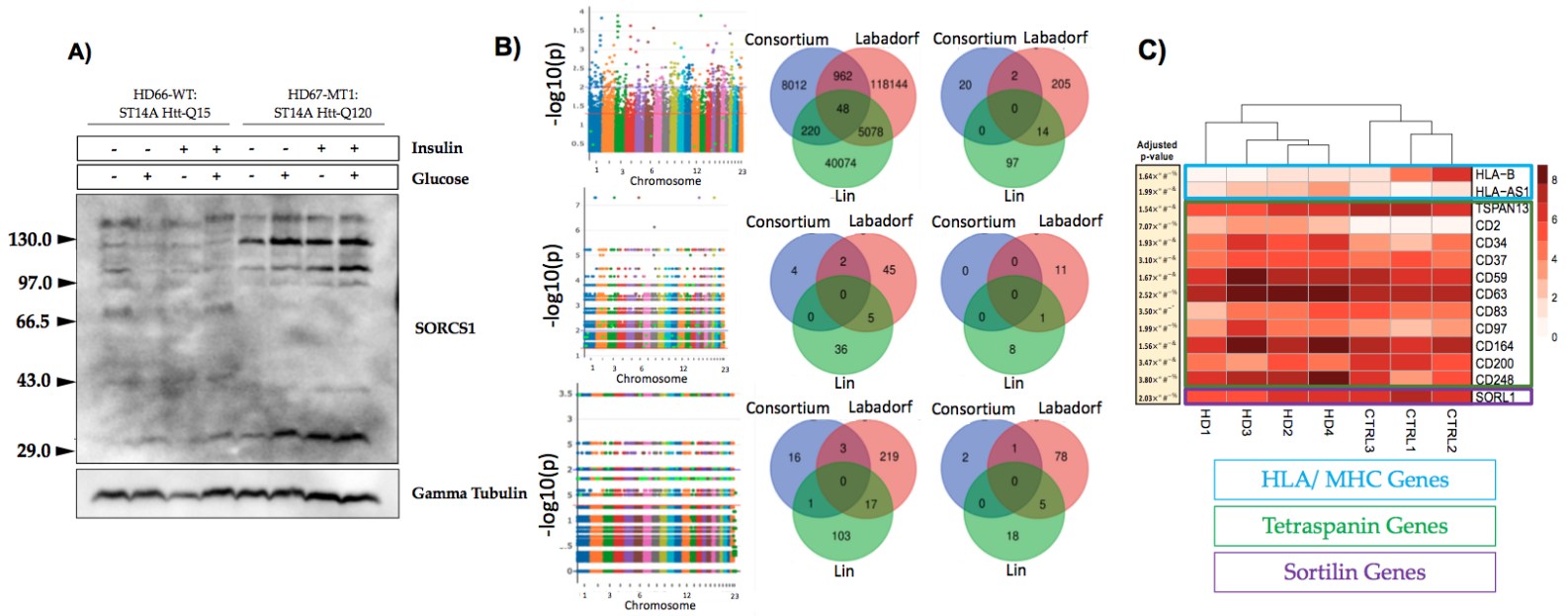

Figure 3.2: Influence of the mutant huntingtin (mHTT) on the expression and genetic variance of sortilin genes in *Rattus norvegicus* (A) and human (B). (A) Western blot of sortilin related VPS10 domain containing receptor 1 (SORCS1) protein in ST14A cells after overnight exposure to a growth medium containing glucose and bovine insulin; (B) Manhattan plots showing SNPs detected in Labdorf, Lin and HD iPSC Consortium datasets (top to bottom). Venn diagrams depict SNPs detected in three human datasets analyzed per sortilin gene (HD iPSC Consortium, Blue; Labadorf 2015, Red; Lin 2016, Green). First the Venn diagram shows SNPs in the three human datasets associated with HD ($p < 0.05$). The other five Venn diagrams show SNPs flanking the five sortilins *SORT1*, *SORL1*, *SORCS1*, *SORCS2* and *SORCS3*); (C) RNA-Seq analysis using the DESeq2 R package on the HD iPSC Consortium dataset. Blue: MHC/HLA; Green: Tetraspanins; Purple: VPS10P (sortilins).

Labadorf *et al.* identified the expression of sortilins affected in HD individuals [110]. We believe that the HD iPSC Consortium dataset did not share common SNPs with the Labadorf and Lin datasets because the HD iPSC Consortium dataset was composed of "HD patient-derived iPSC lines with juvenile-onset CAG repeat expansions (60 and 109 repeats)", a more genetically homogeneous group than the Labadorf and Lin datasets (Figure 2B). Further confirming our observations and those of the Labadorf group gene expression of sortilins are influenced by mHTT in the HD iPSC Consortium dataset (Figure 2C). Supplementary Table 1 shows the description of SNPs that are found significant across all human datasets (p-value < 0.05, Fisher's exact test). Among the biological processes reported on the Supplementary Table 1, tetraspanin variants (illustrated in bold on Supplementary Table S1) indicate the association of mHTT with the four-transmembrane proteins. In the ST14A rat cells that we studied previously and here, the top up-regulated gene was TSPAN8, a transmembrane protein responsible for the organization of the HLA/MHC complex. This observation indicates the agreement between the findings concerning tetraspanins in the human and rat datasets analyzed in this study. To compare the distribution of SNPs identified near the sortilin regions in the GATK/GWAS pipeline exploited in this study with the overall SNPs flanking other genes, Manhattan plots were constructed with the significant (p-value < 0.05) 84,349 SNPs (Table 1) detected in the Labadorf dataset as shown in Figure 3.

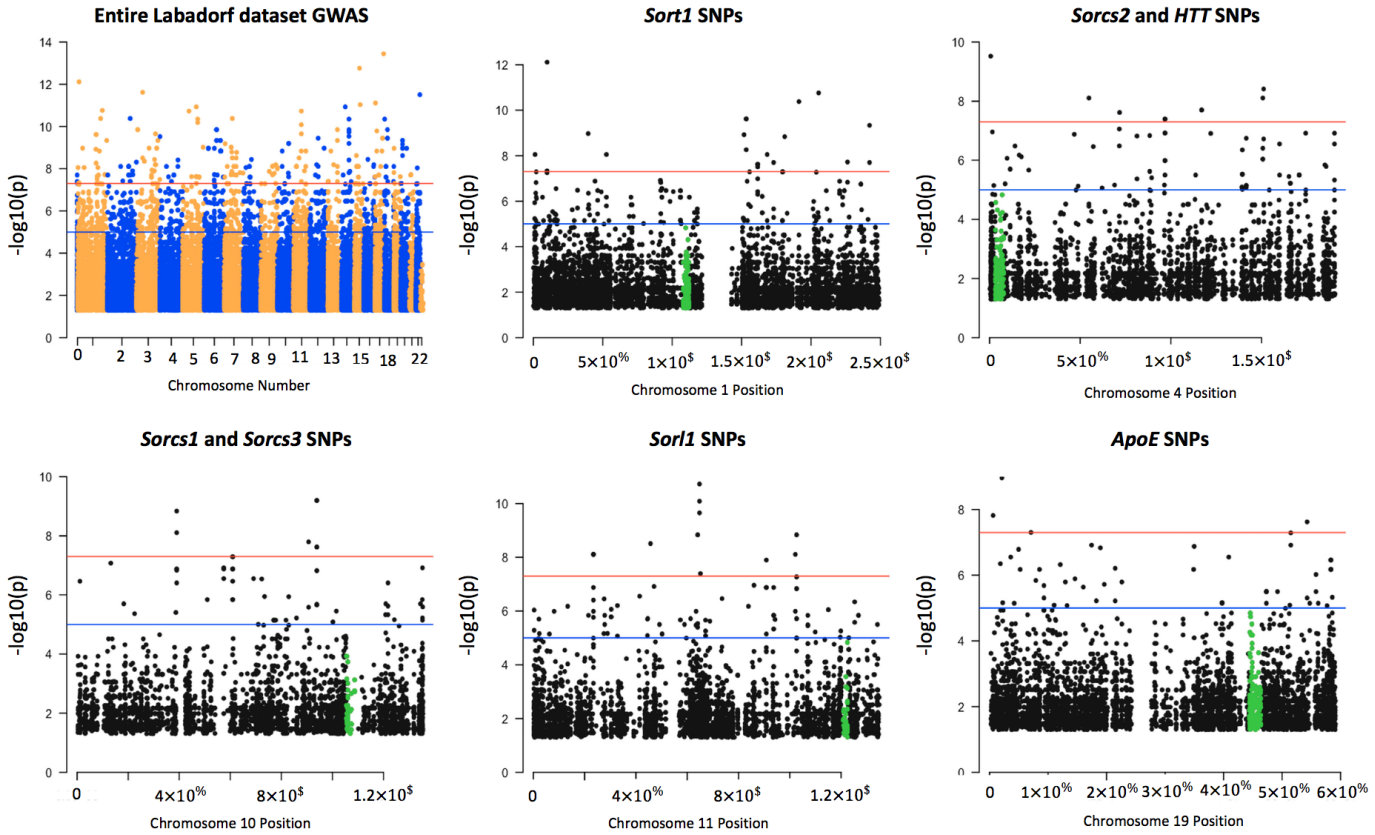

Figure 3.3: Manhattan plot visualization of SNPs found significant (p-value < 0.05) in regions near *HTT*, sortilins and *APOE* in the Labadorf 2015 dataset. Some of the genes were located in the same chromosome (*HTT* and *SORCS2* on chromosome 4, *SORCS1* and *SORCS3* on chromosome 10). When the chromosome was identical, significant SNPs were counted as SNPs located 1,000,000 bp upstream and downstream the gene border coordinates. The level of significance (negative log of p-value, y axis) of SNPs identified near the sortilin genes suggested that the genetic variation in sortilins have a significant impact on the Huntington's disease (HD) pathology in this dataset.

We included visualization of the *APOE* gene variants in Figure 3 due to its importance in neurodegenerations, Alzheimer's Disease in particular. *APOE* genetic variants also affect human longevity due to the transport of lipids and risks associated with cardiovascular diseases [122]. Total SNPs detected between HTT and *SORCS2* on chromosome 4 (11746 SNPs, Supplementary Materials Table S6) were approximately 5X more SNPs than the region between SORCS1 and *SORCS3* genes on chromosome 10 (1976 SNPs, Supplementary Materials Table S6). This is in agreement with the notion of the biological importance of the *HTT* locus as the Mendelian causative mutation of the HD pathology. Compared to the *APOE* locus which is highly polymorphic due to its activity in lipid metabolism, *HTT* carried at least 5X more SNP mutations than *APOE* (*APOE* carried 1863 total SNPs in its vicinity) (Figure 3, Supplementary Materials Table S6). We detected 665 SNPs significantly associated with HD in the vicinity of the *HTT* and sortilins genes (Supplementary Materials Table 6). Positions, alleles, allele frequencies and p-values of 665 SNPs between the *HTT* and sortilins genes significantly associated with HD are indicated in Supplementary Materials Table S7.

3.3.2 mHTT Affects Pathways Important for Immunological Function

By interrogating datasets analyzed for transcriptomic regulation and presence of SNPs associated with mHTT in specific genomic regions known to be involved with Type 1 and Type 2 Diabetes, we found that genes from the MHC/HLA locus, regarded as T1D markers [18], were down-regulated in human cells expressing mHTT in the HD iPSC Consortium dataset (*HLA-B*) (Figure 2C). MHC/HLA genes are responsible

for the presentation of processed antigens to T-cells, and the activation of adaptive immune response, as the gene expression is turned on after a virus or bacterial infection or tumor mutation of proteins [7]. Figure 2C also shows that sortilin *SORL1* and several tetraspanins are modulated by HD in the HD iPSC Consortium dataset. *SORL1* gene expression was down-regulated to approximately a third of the control values, and reached statistical significance (Supplementary Table S3). Other sortilins showed a similar trend, but not significance in the RNA-Seq analysis (Supplementary Table S3). The *HLA-B* gene showed a decrease in the gene expression average of more than 50X in the HD iPSC Consortium dataset (Supplementary Table S3). The trend in the expression of the *SORL1* gene in human was opposite to the pattern of the protein expression of the *SORCS1* gene observed in ST14 rat cells, in that the *SORCS1* protein was up-regulated by almost 2X (for the quantification of protein bands shown on Figure 2A, see Supplementary Materials Figure S2 and Supplementary Table S5) in ST14A cells. These observations are based on the R package DESeq2 analysis [55] that we performed on the datasets, which was different than that performed by Labadorf et al. The *SORCS1* gene in ST14A cells was also up-regulated in the DESeq2 RNA-Seq analysis (Supplementary Table S4). In human datasets, sortilins showed a trend to be down-regulated in the HD iPSC Consortium dataset (Figure 2C, Supplementary Table S3), in the Labadorf dataset (Supplementary Table S6) and in the Lin 2016 dataset (Supplementary Table S7). However, the DESeq2 analysis was only detected as statistically significant, the difference in the expression levels of the *SORL1* gene in the HD iPSC Consortium dataset (Figure 2C), suggesting a gene expression influence of

mHTT on *SORL1* similar to the trend detected previously in the AD for this receptor of Apolipoprotein E [123, 124]. These findings suggest a need in characterization of gene expression alterations related to the immune function of sortilins in HD.

3.3.3 Validation of GATK/GWAS DNA Variants Identification by Pyrosequencing Methodology

To validate the GATK/GWAS pipeline and aiming to establish the presence of an SNP as the result of RNA or DNA mutation, we performed a pyrosequencing detection of SORCS1 SNPs in DNA extracted from ST14A cells carrying HTT and mHTT. To assess whether reads spanning SNPs loci in mutant cells were the result of cell population heterogeneity or clonal expansion of the SNP, we analyzed single cell data previously acquired in our lab from the same ST14A cells. Figure 4A shows the Integrative Genome Viewer (IGV) visualization of reads spanning the loci of the detected SNP. Figure 4B shows pyrosequencing results with mutation identified in genomic DNA. Since the T allele was present in both the IGV visualization and DNA samples interrogated by pyrosequencing, we conclude that the variant was a DNA variant rather than a SORCS1 mRNA variant. This result confirmed that this SNP mutation in SORCS1 was identical to the mutation detected by the GATK/GWAS pipeline using ST14A mRNA molecules, a C > T DNA substitution (Figure 2.4B).

Figure 3.4: Visualization of mutations detected in the GATK/GWAS pipeline and validation of the pipeline by genomic DNA sequencing of samples using pyrosequencing. (A) Genome browser visualization of single-cell RNA-Seq reads spanning the *SORCS1* gene in mutant (mHTT) and wild-type cells. The figure shows mutant samples with zero, five and six reads spanning the *SORCS1* locus; (B) confirmation of detection of C>T substitution in the DNA sequence by pyrosequencing in *SORCS1*.

3.3.4 SNPs in Sortilins and *HTT* identified by GATK/GWAS Pipeline in Linkage Disequilibrium

We interrogated 166 SNPs reported in other studies about their presence in the Labadorf dataset. The Supplementary Table S2 shows SNPs identified in this GWAS study. 48 SNPs were present in the Labadorf dataset, indicating a detection rate close to 30% of well-documented publicly available sortilins SNPs in our study. Not necessarily all 48 SNPs had a significant p-value suggestive of association with HD (Supplementary Table S2). However, important genetic variants previously detected in studies associated with Alzheimer's Disease were associated with HD (Supplementary Table S2, p-values $<$ 0.05). Results shown in Supplementary Table S2, although associated with HD (those with p-values $<$ 0.05) revealed that few sortilins SNP variants were in Linkage with HD (Supplementary Material Figure S3). Therefore, we investigated Linkage patterns unique to our GATK/GWAS dataset using the 665 SNPs shown in Figure 3 as associated with HD. A complete list of SNP variants per gene and identity of the 665 significant SNPs in HTT and sortilins genes are shown in Supplementary Materials Tables S8 and S9. After manually checking for the correlation of SNP position inheritance in the control and cases of HD, haplotype maps of the Linkage Disequilibrium analysis involving *HTT*, *SORT1*, *SORL1*, *SORCS1*, 2 and 3 SNPs variants are shown in Figure 5. Linkage patterns observed in the Labadorf dataset on Figure 5 were in agreement with the Linkage pattern observed in Lin 2016 (Supplementary Materials Figure S4). In both datasets, a continuous haplotype block including SNP Chr1:109950858 (indicated in the

red square in Figure 5), upstream of SORT1 is in association with another haplotype block in region Chr4:3236883-3238643 near HTT in HD cases, but not in controls of both Labadorf and Lin datasets (Figure 5, Supplementary Materials Figure S4). Genes, positions and p-values of SNPs associated with HD shown in Figure 5 are indicated in Supplementary Materials Table S10.

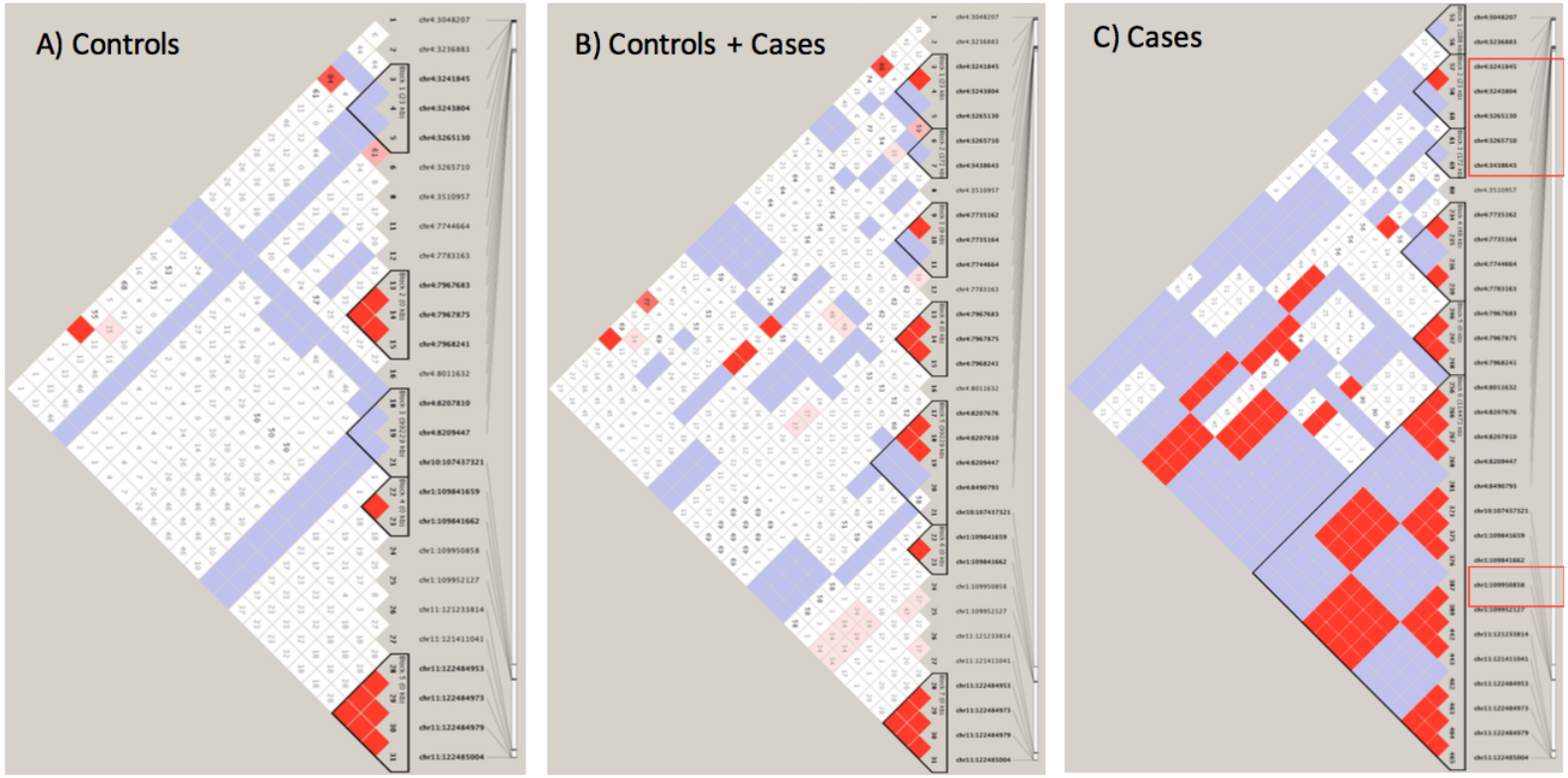

Figure 3.5: Linkage Disequilibrium of sortilin variants with *HTT* in human cases of the Labadorf dataset. Coordinates in chromosome 4 represent SNPs in the *HTT* and *SORCS2*. Coordinates in chromosome 10 represent SNPs in *SORCS1* and *SORCS3*. Coordinates in chromosome 1 and 11 represent SNPs in *SORT1* and *SORL1*, respectively. (A) Genotypes of control individuals; (B) genotypes of the control and cases combined; (C) genotypes of cases. High Linkage indicated by the shades of red between the *HTT* (position 3–3.2M of chromosome 4) and all sortilins.

3.4 Discussion

Previously, we reported on the modulation in the expression of glucose transporters GLUT1 and GLUT4, and sortilin SORCS1 by mHTT in the ST14A cells, suggesting that mechanisms of insulin release and insulin sensitivity were affected in human HD by means of sortilin genes [28]. The identification of these genes in the HD pathology may be useful for therapeutic interventions that may include the CRISPR/Cas9 gene editing or small drug inhibition of such genes. Sortilin SORCS1 was described as a potential target in obesity. Sortilins are responsible for protein-protein interaction, internalization and sorting of proteins between the trans-Golgi network and endosomes using the retromer complex [14]. Given this role in cells, we hypothesized that sortilin plays a role in the release of exosome vesicles. To evaluate biological mechanisms of SORCS1, we are currently building a knock-out model of the *SORCS1* ST14A cells using CRISPR/Cas9 gene editing. We found non-coding variants in introns 1 and 2 of sortilin genes, including SORCS1, associated with HD in the human and in rat (Figure 1). This observation suggests that variants could affect the interaction of sortilins genomic sequence with proteins involved in their expression (Figures 1–3 and 5). These results are in agreement with findings of Reitz et al., because these authors report on genetic variants located in introns 1 and 2 of *SORCS1*, associated with Alzheimer's Disease [8]. We also found a total of 11 SNPs, three of which are significantly associated with HD, in common with SNPs observed by Reitz et al. 2013. All SNPs were found in the non-coding regions of sortilins, (significant SNPs, p-value ¡ 0.05: *SORL1*,

Chr11:121439665; *SORCS1*, Chr10:108706022 and *SORCS2*, Chr4:7733843) (Figure 1). *SORCS1* mutations located in intron 2, between exons that encode the VPS10P sorting domain (Figure 1). These SNPs are therefore common to HD and AD, suggesting that sortilins play biological roles in both AD and HD (Figure 1). We are confident about these results because the number of significant SNPs associated with HD was small compared to the total number of SNPs identified by the present GATK/GWAS pipeline (less than 5% in all datasets as shown in Table 1). Furthermore, we confirmed that the mutations detected by this GATK/GWAS pipeline were present in the genomic DNA derived from the ST14A cells using the pyrosequencing assay, ruling out the possibility of mRNA variation (Figure 4A). The Labadorf group also reported modulation of sortilin *SORCS3* by mHTT in their Supplementary Materials [28]. We observed that protein levels of sortilin *SORCS1* were up-regulated in the ST14A cells (Figure 2A), suggesting modification in the normal subcellular traffic between intracellular organelles and plasma-membrane, as expected from cellular roles of sortilins [14, 21]. Vázquez reported on the involvement of sortilins in the maturation of phagosomes in the process of phagocytosis of *Mycobacterium tuberculosis* [21]. In order to display peptides derived from this parasite in the MHC/HLA class II complex, endosomal and lysosomal enzymes need to be delivered in sequence (sorted) into the phagosome which contains the parasite proteins that will ultimately generate the parasite antigens [21]. Adaptor proteins are also implicated in the receptor-ligand transport of sortilins via clathrin-coated vesicles to endosomes. Once cargo is released, induced by the low-pH environment [125], the retrograde transport of sortilin from endosome to trans-Golgi network depends on

the sortilin interaction with the retromer complex [14, 21]. Therefore, results shown in Figures 1–3 and 5 suggest that mHTT affects the expression or associate with genetic variants of sortilins in ST14A cells and human individuals. This is significant as it establishes a direct connection between vesicle tra c towards phagosome maturation for innate and adaptive immune responses and variants of sortilins involved in the process (Figure 3, Figure 5 and Supplementary Table S3). A further investigation of sortilins, tetraspanins and HLA/MHC genes, may lead to the understanding of the involvement of HTT and mHTT in the release of cellular vesicles, including exosomes, and how this might affect immunological pathways in diseases such as HD, diabetes and cancer. One study reported an association of celiac disease (CE), a condition thought to primarily involve MHC/HLA genes, with SORCS1 variants [126]. It is not surprising to find vesicle tra cking associated with defects in the HLA/MHC loci from the perspective of sortilins affecting vesicle tra cking in HD brought by our study. Before identification of mHTT as the cause of HD, one hypothesis was that the HLA/MHC loci were potential contributors or the only cause of HD. Consistent with this hypothesis, attempts were performed to assess the involvement of the HLA/MHC locus in the etiology of the Huntington's Disease [127, 26]. Although some association was initially reported, it did not hold true after scrutiny of statistical correction [127]. We observed several of the HLA/MHC rat genes with their expression influenced by mHTT in ST14A cells (data not shown). We are currently investigating the involvement of MHC/HLA genes in our ST14A HD cell model. Our results add to the growing body of knowledge of sortilins as proteins that affect processing of antigens in autoimmunity and other functions related

to the immune system by providing evidence that gene expression or genetic diversity of sortilin and MHC/HLA genes are associated with mHTT in HD (Figures 1–3 and 5; Supplementary Materials Table S3). These findings contribute to an improved understanding on the control of tra c of glucose receptors, insulin and Alzheimer's Disease proteins in the trans-Golgi network, the endosome and the plasma-membrane in normal and HD conditions, using models of HD described in this report (Figures 1–5). We showed that mHTT is associated with increased protein expression of sortilin *SORCS1* in ST14A cells (Figure 2A), in agreement with our previous report for mRNA molecules of *SORCS1* and that there exists allele association (also known as Linkage Disequilibrium) of regions near *SORT1*, *SORL1*, *SORCS1*, *SORCS2* and *SORCS3* with gene variants in the region of *HTT* gene specifically in human HD cases (Figure 5). Eleven sortilins SNPs previously detected by Reitz et al. (2013), who investigated the involvement of *SORCS1* variants with AD, were detected in our study (Supplementary Table S2) [14, 8, 15]. It is important to contrast the number of individuals involved in many GWAS studies to the number of HD cases reported here (Reitz 2013 individuals: n = 11,840 cases and 10,931 controls; Labadorf dataset 2015: n = 20 cases and 49 controls). Due to the sample size, more genetic variability is associated with the design of the Reitz group, which helps explain the low number of positions identified in common between the Labadorf and Reitz studies. Some of the SNP variants associated with HD were reported in Alzheimer's and cardiovascular diseases GWASs (Supplementary Table S2).

3.5 Conclusions

We present evidence of the association between the mHTT and genes of critical biological processes in neurodegenerative diseases, diabetes and cancer in rat and human RNA-Seq datasets. The genetic association of VPS10P (sortilins) variants with HTT variants in HD cases has also been established in this work. Genetic variance of the HLA/MHC loci in human and rat samples deserve further investigation, given the association with the mHTT shown in human and rat datasets. We hypothesize the existence of an allele association between sortilins and HLA/MHC genes. We envision sortilins and HLA/MHC proteins as potential therapeutic targets in cancer, diabetes, and neurodegenerative diseases, by modulation of antigen processing and immune response mechanisms in these diseases. In addition, our results further validate the rat ST14A cells as a model of human HD. An overview of these mechanisms related to cellular immune functions can be identified in Figure 25.

Chapter 4

Concluding Remarks

We began with quantification of glucose levels in single ST14A cells, a model resembling the human striatum affected by Huntington's disease, using the nanopipette platform. Glucose levels were regulated by mHTT, suggesting that phenotypes of diabetes previously observed in human patients could be further investigated and characterized at the molecular level in the cell model used in our laboratory. When we interrogated the same ST14A cells regarding the set of genes whose expression was affected by the mHTT protein, we discovered that several genes present in the glycolysis pathway, pentose phosphate pathway, Krebs cycle, and glucose transporters, all of which are signaling pathways responsible for glucose homeostasis in the cell, were regulated by mHTT protein. This finding illustrates that intracellular levels of glucose were actually controlled at the genomic and proteomic levels by the expression of the genes involved in the above-mentioned signaling pathways. To our surprise, the genes we chose for further molecular biology validation, sortilin *SORCS1* and glucose receptors, supported

very well the hypothesis that type 1 and type 2 diabetes phenotypes were present in the ST14A cells expressing the mutant huntingtin protein (mHTT) (Chapter 2). We showed that sortilins could well be involved in Type 2 diabetes phenotypes in HD models, as supported by their role in the translocation of GLUT4 receptors, down-regulated in mutant ST14A cells by mHTT. In Chapter 3, we investigated the involvement of genes identified by the RNA-Seq analysis of ST14A cells in human samples. We found SNPs as well as gene expression of sortilins to be associated with human HD. Surprisingly, we found that sortilins have an immunological role in cells by participating in phagosome maturation and the display of class II MHC peptides in macrophages. This piece of evidence described in the literature, agreed well with the molecular function of the top upregulated gene, *TSPAN8*, a tetraspanin. It has been proposed that tetraspanins help organize a lipid domain in the cellular membrane, also composed of MHC proteins. MHC proteins may be translocated to the cellular membrane by the cellular machinery composed of vesicles transported using the cytoskeleton and HTT coupled to the molecular motor complex. Taken together, these findings suggest that diabetes phenotypes in Huntington's disease are inherited through the machinery associated with the subcellular traffic of glucose transporters and MHC proteins. In contrast to previous hypotheses, our results suggest that these subcellular traffic mechanisms likely involve sortilin proteins in both Type 1 and Type 2 diabetes phenotypes.

Bibliography

[1] K. B. Bhattacharyya, "The story of george huntington and his disease," *Ann Indian Acad Neurol*, vol. 19, no. 1, pp. 25–8, 2016.

[2] F. Saudou and S. Humbert, "The biology of huntingtin," *Neuron*, vol. 89, no. 5, pp. 910–26, 2016.

[3] H. E. Wu, T. Melicher, B. Cao, M. Sanches, J. A. Stanley, G. Zunta-Soares, B. Mwangi, and J. C. Soares, "Mr spectroscopy findings of the basal ganglia in bipolar disorders: a systematic review," *Current Psychiatry Reviews*, vol. 14, 2016.

[4] S. Ferré, J. Bonaventura, W. Zhu, C. Hatcher-Solis, J. Taura, C. Quiroz, N. S. Cai, E. Moreno, V. Casadó-Anguera, A. V. Kravitz, K. R. Thompson, D. G. Tomasi, G. Navarro, A. Cordomí, L. Pardo, C. Lluís, C. W. Dessauer, N. D. Volkow, V. Casadó, F. Ciruela, D. E. Logothetis, and D. Zwilling, "Essential control of the function of the striatopallidal neuron by pre-coupled complexes of adenosine a," *Front Pharmacol*, vol. 9, p. 243, 2018.

[5] F. Gardoni and C. Bellone, "Modulation of the glutamatergic transmission by

dopamine: a focus on parkinson, huntington and addiction diseases," *Front Cell Neurosci*, vol. 9, p. 25, 2015.

[6] J. C. Hou and J. E. Pessin, "Ins (endocytosis) and outs (exocytosis) of glut4 trafficking," *Curr Opin Cell Biol*, vol. 19, no. 4, pp. 466–73, 2007.

[7] A. Abbas, A. Lichtman, and S. Pillai, *Cellular and Molecular Immunology*. ELSEVIER, 2017.

[8] C. Reitz, G. Tosto, B. Vardarajan, E. Rogaeva, M. Ghani, R. S. Rogers, C. Conrad, J. L. Haines, M. A. Pericak-Vance, M. D. Fallin, T. Foroud, L. A. Farrer, G. D. Schellenberg, P. S. George-Hyslop, R. Mayeux, and A. D. G. C. (ADGC), "Independent and epistatic effects of variants in vps10-d receptors on alzheimer disease risk and processing of the amyloid precursor protein (app)," *Transl Psychiatry*, vol. 3, p. e256, 2013.

[9] R. Adam Seger, P. Actis, C. Penfold, M. Maalouf, B. Vilozny, and N. Pourmand, "Voltage controlled nano-injection system for single-cell surgery," *Nanoscale*, vol. 4, no. 19, pp. 5843–6, 2012.

[10] M. T. Montojo, M. Aganzo, and N. González, "Huntington's disease and diabetes: Chronological sequence of its association," *J Huntingtons Dis*, vol. 6, no. 3, pp. 179–188, 2017.

[11] J. F. Gusella, N. S. Wexler, P. M. Conneally, S. L. Naylor, M. A. Anderson, R. E. Tanzi, P. C. Watkins, K. Ottina, M. R. Wallace, and A. Y. Sakaguchi,

"A polymorphic dna marker genetically linked to huntington's disease," *Nature*, vol. 306, no. 5940, pp. 234–8, 1983.

[12] T. H. D. C. R. Group, "A novel gene containing a trinucleotide repeat that is expanded and unstable on huntington's disease chromosomes," *Cell*, vol. 72, no. 6, pp. 971–83, 1993.

[13] E. Cattaneo and L. Conti, "Generation and characterization of embryonic striatal conditionally immortalized st14a cells," *J Neurosci Res*, vol. 53, no. 2, pp. 223–34, 1998.

[14] C. Reitz, "The role of the retromer complex in aging-related neurodegeneration: a molecular and genomic review," *Mol Genet Genomics*, vol. 290, no. 2, pp. 413–27, 2015.

[15] C. Reitz, S. Tokuhiro, L. N. Clark, C. Conrad, J. P. Vonsattel, L. N. Hazrati, A. Palotás, R. Lantigua, M. Medrano, I. Z Jiménez-Velázquez, B. Vardarajan, I. Simkin, J. L. Haines, M. A. Pericak-Vance, L. A. Farrer, J. H. Lee, E. Rogaeva, P. S. George-Hyslop, and R. Mayeux, "Sorcs1 alters amyloid precursor protein processing and variants may increase alzheimer's disease risk," *Ann Neurol*, vol. 69, no. 1, pp. 47–64, 2011.

[16] R. A. Nascimento, R. E. Özel, W. H. Mak, M. Mulato, B. Singaram, and N. Pourmand, "Single cell "glucose nanosensor" verifies elevated glucose levels in individual cancer cells," *Nano Lett*, vol. 16, no. 2, pp. 1194–200, 2016.

[17] G. Bulbul, G. Chaves, J. Olivier, R. E. Ozel, and N. Pourmand, "Nanopipettes as monitoring probes for the single living cell: State of the art and future directions in molecular biology," *Cells*, vol. 7, no. 6, 2018.

[18] S. Nejentsev, J. M. Howson, N. M. Walker, J. Szeszko, S. F. Field, H. E. Stevens, P. Reynolds, M. Hardy, E. King, J. Masters, J. Hulme, L. M. Maier, D. Smyth, R. Bailey, J. D. Cooper, G. Ribas, R. D. Campbell, D. G. Clayton, J. A. Todd, and W. T. C. C. Consortium, "Localization of type 1 diabetes susceptibility to the mhc class i genes hla-b and hla-a," *Nature*, vol. 450, no. 7171, pp. 887–92, 2007.

[19] K. W. Wucherpfennig and D. Sethi, "T cell receptor recognition of self and foreign antigens in the induction of autoimmunity," *Semin Immunol*, vol. 23, no. 2, pp. 84–91, 2011.

[20] M. A. Kebede, A. T. Oler, T. Gregg, A. J. Balloon, A. Johnson, K. Mitok, M. Rabaglia, K. Schueler, D. Stapleton, C. Thorstenson, L. Wrighton, B. J. Floyd, O. Richards, S. Raines, K. Eliceiri, N. G. Seidah, C. Rhodes, M. P. Keller, J. L. Coon, A. Audhya, and A. D. Attie, "Sorcs1 is necessary for normal insulin secretory granule biogenesis in metabolically stressed β cells," *J Clin Invest*, vol. 124, no. 10, pp. 4240–56, 2014.

[21] C. L. Vázquez, A. Rodgers, S. Herbst, S. Coade, A. Gronow, C. A. Guzman, M. S. Wilson, M. Kanzaki, A. Nykjaer, and M. G. Gutierrez, "The proneu-

rotrophin receptor sortilin is required for mycobacterium tuberculosis control by macrophages," *Sci Rep*, vol. 6, p. 29332, 2016.

[22] H. Talbot, S. Saada, T. Naves, P. F. Gallet, A. L. Fauchais, and M. O. Jauberteau, "Regulatory roles of sortilin and sorla in immune-related processes," *Front Pharmacol*, vol. 9, p. 1507, 2018.

[23] P. A. Roche and K. Furuta, "The ins and outs of mhc class ii-mediated antigen processing and presentation," *Nat Rev Immunol*, vol. 15, no. 4, pp. 203–16, 2015.

[24] G. Ellrichmann, C. Reick, C. Saft, and R. A. Linker, "The role of the immune system in huntington's disease," *Clinical and Developmental Immunology*, vol. 2013, pp. 1–11, 2013.

[25] A. Mastrokolias, Y. Ariyurek, J. J. Goeman, E. van Duijn, R. A. Roos, R. C. van der Mast, G. B. van Ommen, J. T. den Dunnen, P. A. 't Hoen, and W. M. van Roon-Mom, "Huntington's disease biomarker progression profile identified by transcriptome sequencing in peripheral blood," *European Journal of Human Genetics*, vol. 23, no. 10, pp. 1349–1356, 2015.

[26] H. Madsen, L. S. Nielsen, and S. A. Sørensen, "An association study of huntington's disease and hla," *J Med Genet*, vol. 19, no. 6, pp. 452–4, 1982.

[27] G. Chaves, J. Stanley, and N. Pourmand, "Mutant huntingtin affects diabetes and alzheimer's markers in human and cell models of huntington's disease," *Cells*, vol. 8, no. 9, 2019.

[28] G. Chaves, R. E. Özel, N. V. Rao, H. Hadiprodjo, Y. D. Costa, Z. Tokuno, and N. Pourmand, "Metabolic and transcriptomic analysis of huntington's disease model reveal changes in intracellular glucose levels and related genes," *Heliyon*, vol. 3, no. 8, p. e00381, 2017.

[29] S. E. Andrew, Y. P. Goldberg, B. Kremer, H. Telenius, J. Theilmann, and S. Adam, "The relationship between trinucleotide (cag) repeat length and clinical features of huntington's disease," *Nat. Genet.*, vol. 4, pp. 398–403, 1993.

[30] C. Zuccato, A. Ciammola, D. Rigamonti, B. R. Leavitt, D. Goffredo, and L. Conti, "Loss of huntingtin-mediated bdnf gene transcription in huntington's disease," *Science*, vol. 293, pp. 493–498, 2001.

[31] E. S. Nambron, E. Kalliolia, C. Ottolenghi, P. Hindmarsh, and N. R. Hill, "et al," *A Metabolic Study of Huntington's Disease, PLoS One*, vol. 11, 2016.

[32] H. Mizuno, H. Shibayama, F. Tanaka, M. Doyu, G. Sobue, and H. Iwata, "An autopsy case with clinically and molecular genetically diagnosed huntington's disease with only minimal non-specific neuropathological findings," *Clin Neuropathol.*, vol. 19, pp. 94–103, 2000.

[33] E. Gómez-Tortosa, M. E. MacDonald, J. C. Friend, S. A. Taylor, L. J. Weiler, and L. A. Cupples, "Quantitative neuropathological changes in presymptomatic huntington's disease," *Ann Neurol.*, vol. 49, pp. 29–34, 2001.

[34] T. Pancani, D. J. Foster, M. S. Moehle, T. J. Bichell, E. Bradley, and T. M.

Bridges, "Allosteric activation of m4 muscarinic receptors improve behavioral and physiological alterations in early symptomatic yac128 mice," in *Proc. Natl. Acad. Sci. USA 112*, pp. 14078–14083, 2015.

[35] J. Modregger, N. A. DiProspero, V. Charles, D. A. Tagle, and M. Plomann, "Pacsin 1 interacts with huntingtin and is absent from synaptic varicosities in presymptomatic huntington's disease brains," *Hum. Mol. Genet*, vol. 11, pp. 2547–2558, 2002.

[36] X. Liang, J. Wu, P. Egorova, and I. Bezprozvanny, "An automated and quantitative method to evaluate progression of striatal pathology in huntington's disease transgenic mice," *J. Huntingtons Dis.*, vol. 3, pp. 343–350, 2014.

[37] A. Labadorf, A. G. Hoss, V. Lagomarsino, J. C. Latourelle, T. C. Hadzi, J. Bregu, M. E. MacDonald, J. F. Gusella, J. F. Chen, S. Akbarian, Z. Weng, and R. H. Myers, "Rna sequence analysis of human huntington disease brain reveals an extensive increase in inflammatory and developmental gene expression," *PLoS One*, vol. 10, no. 12, p. e0143563, 2015.

[38] M. T. Lin and M. F. Beal, "Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases," *Nature*, vol. 443, pp. 787–795, 2006.

[39] A. I. I. Acuña, M. Esparza, C. Kramm, F. A. Beltrán, A. V. Parra, and C. Cepeda, "A failure in energy metabolism and antioxidant uptake precede symptoms of huntington's disease in mice," *Nat. Commun.*, vol. 4, p. 2917, 2013.

[40] W. I. Sivitz, S. L. DeSautel, T. Kayano, G. I. Bell, and J. E. Pessin, "Regulation of glucose transporter messenger rna levels in rat adipose tissue by insulin," *Mol. Endocrinol.*, vol. 4, pp. 583–588, 1990.

[41] E. M. Quistgaard, C. Löw, F. Guettou, and P. Nordlund, "Understanding transport by the major facilitator superfamily (mfs): structures pave the way," *Nat. Rev. Mol. Cell Biol*, vol. 17, pp. 123–132, 2016.

[42] H. Lund-Andersen, "Transport of glucose from blood to brain," *Physiol. Rev.*, vol. 59, pp. 305–352, 1979.

[43] W. M. Pardridge, R. J. Boado, and C. R. Farrell, "Brain-type glucose transporter (glut-1) is selectively localized to the blood-brain barrier; studies with quantitative western blotting and in situ hybridization," *J. Biol. Chem. 265*, vol. 265, pp. 18035–18040, 1990.

[44] S. J. Vannucci, F. Maher, and I. A. Simpson, "Glucose transporter proteins in brain: delivery of glucose to neurons and glia," *Glia*, vol. 21, pp. 2–21, 1997.

[45] I. S. Wood and P. Trayhurn, "Glucose transporters (glut and sglt): expanded families of sugar transport proteins,," *Br. J. Nutr.*, vol. 89, pp. 3–9, 2003.

[46] R. Kaddurah-Daouk and K. R. R. Krishnan, "Metabolomics: a global biochemical approach to the study of central nervous system diseases," *Neuropsychopharmacology*, vol. 34, pp. 173–186, 2009.

[47] R. E. E. Özel, A. Lohith, W. H. Mak, and N. Pourmand, "Single-cell intracellular nano-ph probes," *RSC Adv*, vol. 5, pp. 52436–52443, 2015.

[48] M. W. Pfaffl, "A new mathematical model for relative quantification in real-time rt-pcr," *Nucleic Acids Res*, vol. 29, 2001.

[49] J. Lazar, "Sorcs1 contributes to the development of renal disease in rats and humans," *Physiol. Genomics*, vol. 45, no. 16, pp. 720–728, 2013.

[50] S. Khowal, "Assessment of the therapeutic potential of hesperidin and proteomic resolution of diabetes-mediated neuronal fluctuations expediting alzheimer's disease,," *RSC Adv*, vol. 5, no. 58, pp. 46965–46980, 2015.

[51] C. Chen, S. S. Khaleel, H. Huang, and C. H. Wu, "Software for pre-processing illumina next-generation sequencing short read sequences," in *Source Code Biol Med*, 9 8, 2014.

[52] A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, and S. Jha, "Star: ultrafast universal rna-seq aligner," *Bioinformatics*, vol. 29, pp. 15–21, 2013.

[53] M. I. Love, S. Anders, V. Kim, and W. Huber, "Rna-seq workflow: gene-level exploratory analysis and differential expression," *F1000Res.*, vol. 4, p. 1070, 2015.

[54] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for rna-seq data with deseq2," *Genome Biol*, vol. 15, no. 12, p. 550, 2014.

[55] S. Anders and W. Huber, "Differential expression analysis for sequence count data," *Genome Biol*, vol. 11, 2010.

[56] W. Huber, V. J. Carey, R. Gentleman, S. Anders, M. Carlson, and B. S. Carvalho, "Orchestrating high-throughput genomic analysis with bioconductor," *Nat Methods*, vol. 12, pp. 115–121, 2015.

[57] F. Mochel, P. Charles, F. Seguin, J. Barritault, C. Coussieu, and L. Perin, "Early energy deficit in huntington disease: identification of a plasma biomarker traceable during disease progression," *PLoS One*, vol. 2, 2007.

[58] V. B. Mattis and C. N. Svendsen, *Modeling Huntington's disease with patient-derived neurons.* Brain Res, 2015.

[59] A. Ciarmiello, G. Giovacchini, S. Orobello, L. Bruselli, F. Elifani, and F. Squitieri, "18f-fdg pet uptake in the pre-huntington disease caudate affects the time-to-onset independently of cag expansion size, eur," pp. 1030–1036, J. Nucl. Med. Mol. Imaging 39, 2012.

[60] M. T. Besson, K. Alegría, P. Garrido-Gerter, L. F. Barros, and J.-C. C. Liévens, "Enhanced neuronal glucose transporter expression reveals metabolic choice in a hd drosophila model," *PLoS One*, vol. 10, 2015.

[61] P. Mergenthaler, U. Lindauer, G. A. Dienel, and A. Meisel, "Sugar for the brain: the role of glucose in physiological and pathological brain function," *Trends Neurosci.*, vol. 36, pp. 587–597, 2013.

[62] P. Ebeling, H. A. Koistinen, and V. A. Koivisto, "Insulin-independent glucose transport regulates insulin sensitivity," *FEBS Lett.*, vol. 436, pp. 301–303, 1998.

[63] M. Scott, D. Flaherty, and J. Currall, "Statistics: are we related," *J. Small Anim. Pract.*, vol. 54, pp. 124–128, 2013.

[64] R. C. Team, "R: A language and environment for statistical computing," *Foundation for Statistical Computing, Vienna, Austria*, 2013.

[65] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry.* New York: H. Freeman, 5th ed., 2002.

[66] A. J. Cooper, K. F. Sheu, J. R. Burke, W. J. Strittmatter, and J. P. Blass, "Glyceraldehyde 3-phosphate dehydrogenase abnormality in metabolically stressed huntington disease fibroblasts," *Dev Neurosci.*, vol. 20, pp. 462–468, 1998.

[67] J. C. Corona, A. Gimenez-Cassina, F. Lim, and J. Díaz-Nido, "Hexokinase ii gene transfer protects against neurodegeneration in the rotenone and mptp mouse models of parkinson's disease," *J Neurosci. Res*, vol. 88, pp. 1943–1950, 2010.

[68] D. J. Roberts and S. Miyamoto, "Hexokinase ii integrates energy metabolism and cellular protection: Akting on mitochondria and torcing to autophagy," *Cell Death Differ*, vol. 22, p. 364, 2015.

[69] P. H. Ratz, K. M. Berg, N. H. Urban, and A. S. Miner, "Regulation of smooth muscle calcium sensitivity: Kcl as a calcium-sensitizing stimulus, am," *J. Physiol. Cell Physiol.*, vol. 288, pp. C769–83, 2005.

[70] J. C. Rathmell, C. J. Fox, D. R. Plas, P. S. Hammerman, R. M. Cinalli, and C. B. Thompson, "Akt-directed glucose metabolism can prevent bax conformation change and promote growth factor-independent survival," *Mol. Cell. Biol.*, vol. 23, pp. 7315–7328, 2003.

[71] A. Johri and M. F. Beal, "Antioxidants in huntington's disease," *Biochim. Biophys. Acta*, vol. 1822, pp. 664–674, 2012.

[72] M. Vorgerd and J. Zange, "Carbohydrate oxidation disorders of skeletal muscle," *Curr. Opin. Clin. Nutr. Metab. Care 5*, vol. 5, pp. 611–617, 2002.

[73] M. Ribeiro, T. Rosenstock, A. Oliveira, C. Oliveira, and C. Rego, "Insulin and igf-1 improve mitochondrial function in a pi-3k/akt-dependent manner and reduce mitochondrial generation of reactive oxygen species in huntington's disease knock-in striatal cells," in *Free Radic. Biol. Med.*, pp. 129–144, 2014.

[74] C. M. Taniguchi, B. Emanuelli, and C. R. Kahn, "Critical nodes in signalling pathways: insights into insulin action," *Nat. Rev. Mol. Cell Biol*, vol. 7, pp. 85–96, 2006.

[75] J. N. Savas, L. F. F. Ribeiro, K. D. Wierda, R. Wright, L. A. DeNardo-Wilke, and H. C. Rice, "The sorting receptor sorcs1 regulates trafficking of neurexin and ampa receptors," *Neuron*, vol. 87, pp. 764–780, 2015.

[76] R. F. Lane, P. S. George-Hyslop, B. L. Hempstead, S. A. Small, S. M. Strittmatter,

and S. Gandy, "Vps10 family proteins and the retromer complex in aging-related neurodegeneration and diabetes," *J. Neurosci.*, vol. 32, pp. 14080–14086, 2012.

[77] J. V. Larsen, G. Hermey, E. S. Sørensen, T. Prabakaran, E. I. Christensen, and J. Gliemann, "Human sorcs1 binds sortilin and hampers its cellular functions," *Biochem. J.*, vol. 457, pp. 277–288, 2014.

[78] R. F. Lane, J. W. Steele, D. Cai, M. E. Ehrlich, A. D. Attie, and S. Gandy, "Protein sorting motifs in the cytoplasmic tail of sorcs1 control generation of alzheimer's amyloid-β peptide," *J. Neurosci.*, vol. 33, pp. 7099–7107, 2013.

[79] S. Sekar, J. McDonald, L. Cuyugan, J. Aldrich, A. Kurdoglu, and J. Adkins, "Alzheimer's disease is associated with altered expression of genes involved in immune response and mitochondrial processes in astrocytes," *Neurobiol Aging*, vol. 36, pp. 583–591, 2015.

[80] T. Seredenina and R. Luthi-Carter, "What have we learned from gene expression profiles in huntington's disease?," *Neurobiol. Dis*, vol. 45, pp. 83–98, 2012.

[81] C. Tourette, B. Li, R. Bell, S. O'Hare, L. S. Kaltenbach, and S. D. Mooney, "A large scale huntingtin protein interaction network implicates rho gtpase signaling pathways in huntington disease," *J. Biol. Chem*, vol. 289, pp. 6709–6726, 2014.

[82] C. Cepeda, M. A. Ariano, C. R. Calvert, J. Flores-Hernández, S. H. Chandler, and B. R. Leavitt, "Nmda receptor function in mouse models of huntington disease," *J. Neurosci. Res*, vol. 66, pp. 525–539, 2001.

[83] R. Luthi-Carter, A. Strand, N. L. Peters, S. M. Solano, Z. R. Hollingsworth, and A. S. Menon, "Decreased expression of striatal signaling genes in a mouse model of huntington's disease," *Hum. Mol. Genet*, vol. 9, pp. 1259–1271, 2000.

[84] M. Achour, S. L. Gras, C. Keime, F. Parmentier, F.-X. X. Lejeune, and A.-L. L. Boutillier, "Neuronal identity genes regulated by super-enhancers are preferentially down-regulated in the striatum of huntington's disease mice," *Hum. Mol. Genet*, vol. 24, pp. 3481–3496, 2015.

[85] H.-F. F. Wang, J.-T. T. Yu, W. Zhang, W. Wang, Q.-Y. Y. Liu, and X.-Y. Y. Ma, "Sorcs1 and apoe polymorphisms interact to confer risk for late-onset alzheimer's disease in a northern han chinese population," *Brain Res. 1448*, vol. 1, pp. 111–116, 2012.

[86] A. Nykjaer and T. E. Willnow, "Sortilin: a receptor to regulate neuronal viability and function," *Trends Neurosci.*, vol. 35, pp. 261–270, 2012.

[87] B. Boyuk, S. Degirmencioglu, H. Atalay, S. Guzel, A. Acar, and A. Celebi, "Relationship between levels of brain-derived neurotrophic factor and metabolic parameters in patients with type 2 diabetes mellitus," *J. Diabetes Res.*, vol. 2014, no. 97814, p. 3, 2014.

[88] G. R. Lewin and B. D. Carter, "Neurotrophic factors," *Exp. Pharmacol.*, vol. 220, 2014.

[89] V. Leoni and C. Caccia, "The impairment of cholesterol metabolism in huntington disease," *Biochim. Biophys. Acta*, vol. 1851, pp. 1095–1105, 2015.

[90] A. Strong, Q. Ding, A. C. Edmondson, J. S. Millar, K. V. Sachs, and X. Li, "Hepatic sortilin regulates both apolipoprotein b secretion and ldl catabolism," *J. Clin. Invest*, vol. 122, pp. 2807–2816, 2012.

[91] A. Strong and D. J. Rader, "Sortilin as a regulator of lipoprotein metabolism," *Curr. Atheroscler. Rep*, vol. 14, pp. 211–218, 2012.

[92] D. M. Holtzman, J. Herz, and G. Bu, "Apolipoprotein e receptors: normal biology and roles in alzheimer disease," in *Cold Spring Harb Perspect. Med.*, 2012.

[93] M. Valenza, D. Rigamonti, D. Goffredo, C. Zuccato, S. Fenu, and L. Jamot, "Dysfunction of the cholesterol biosynthetic pathway in huntington's disease," *J. Neurosci.*, vol. 25, pp. 9932–9939, 2005.

[94] A. J. Morton, M. J. Hunt, A. K. Hodges, P. D. Lewis, A. J. Redfern, and S. B. Dunnett, "A combination drug therapy improves cognition and reverses gene expression changes in a mouse model of huntington's disease," *Eur. J. Neurosci.*, vol. 21, pp. 855–870, 2005.

[95] L. A. Farrer, "Diabetes mellitus in huntington disease," *Clin Genet*, vol. 27, no. 1, pp. 62–7, 1985.

[96] K. A. Tobin, H. H. Steineger, S. Alberti, O. Spydevold, J. Auwerx, and J. A.

Gustafsson, "Cross-talk between fatty acid and cholesterol metabolism mediated by liver x receptor-alpha," *Mol. Endocrinol.*, vol. 14, pp. 741–752, 2000.

[97] M. Valenza and E. Cattaneo, "Cholesterol dysfunction in neurodegenerative diseases: is huntington's disease in the list," *Prog. Neurobiol.*, vol. 80, pp. 165–176, 2006.

[98] M. T. Nakamura, B. E. Yudell, and J. J. Loor, "Regulation of energy metabolism by long-chain fatty acids," *Prog. Lipid Res*, vol. 53, pp. 124–144, 2014.

[99] E. J. Stephenson and J. A. Hawley, "Mitochondrial function in metabolic health: a genetic and environmental tug of war," *Biochim. Biophys. Acta*, vol. 1840, pp. 1285–1294, 2014.

[100] G. A. Palidwor, S. Shcherbinin, M. R. Huska, T. Rasko, U. Stelzl, A. Arumughan, R. Foulle, P. Porras, L. Sanchez-Pulido, E. E. Wanker, and M. A. Andrade-Navarro, "Detection of alpha-rod protein repeats using a neural network and application to huntingtin," *PLOS Computational Biology*, vol. 5, pp. 1–11, 03 2009.

[101] S. C. Warby, A. Montpetit, A. R. Hayden, J. B. Carroll, S. L. Butland, H. Visscher, J. A. Collins, A. Semaka, T. J. Hudson, and M. R. Hayden, "Cag expansion in the huntington disease gene is associated with a specific and targetable predisposing haplogroup," *Am J Hum Genet*, vol. 84, no. 3, pp. 351–66, 2009.

[102] S. Podolsky, N. A. Leopold, and D. S. Sax, "Increased frequency of diabetes

mellitus in patients with huntington's chorea," *Lancet*, vol. 1, no. 7765, pp. 1356–8, 1972.

[103] D. Zala, M.-V. Hinckelmann, H. Yu, M. M. Lyra da Cunha, G. Liot, F. P. Cordelières, S. Marco, and F. Saudou, "Vesicular glycolysis provides on-board energy for fast axonal transport," *Cell*, vol. 152, pp. 479–491, 2020/06/10 2013.

[104] K. A. McLaughlin, C. C. Richardson, A. Ravishankar, C. Brigatti, D. Liberati, V. Lampasona, L. Piemonti, D. Morgan, R. G. Feltbower, and M. R. Christie, "Identification of tetraspanin-7 as a target of autoantibodies in type 1 diabetes," *Diabetes*, vol. 65, no. 6, pp. 1690–8, 2016.

[105] E. G. Marcusson, B. F. Horazdovsky, J. L. Cereghino, E. Gharakhanian, and S. D. Emr, "The sorting receptor for yeast vacuolar carboxypeptidase y is encoded by the vps10 gene," *Cell*, vol. 77, no. 4, pp. 579–86, 1994.

[106] E. Reuter, J. Weber, M. Paterka, R. Ploen, T. Breiderhoff, J. van Horssen, T. E. Willnow, V. Siffrin, and F. Zipp, "Role of sortilin in models of autoimmune neuroinflammation," *J Immunol*, vol. 195, no. 12, pp. 5762–9, 2015.

[107] S. Picelli, O. R. Faridani, A. K. Björklund, G. Winberg, S. Sagasser, and R. Sandberg, "Full-length rna-seq from single cells using smart-seq2," *Nat Protoc.*, vol. 9, pp. 171–181, 2014.

[108] Nvigen, "Nvigen size selection kit. available online: www.nvigen.com," 2019.

[109] M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl,

A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, and M. J. Daly, "A framework for variation discovery and genotyping using next-generation dna sequencing data," *Nat Genet*, vol. 43, no. 5, pp. 491–8, 2011.

[110] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo, "The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data," *Genome Res*, vol. 20, no. 9, pp. 1297–303, 2010.

[111] G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo, "From fastq data to high-confidence variant calls: The genome analysis toolkit best practices pipeline," *Current Protocols in Bioinformatics*, vol. 43, no. 1, pp. 11.10.1–11.10.33, 2013.

[112] P. M. Visscher, N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang, "10 years of gwas discovery: Biology, function, and translation," *Am J Hum Genet*, vol. 101, no. 1, pp. 5–22, 2017.

[113] L. Lin, J. W. Park, S. Ramachandran, Y. Zhang, Y. T. Tseng, S. Shen, H. J. Waldvogel, M. A. Curtis, R. L. Faull, J. C. Troncoso, O. Pletnikova, C. A. Ross, B. L. Davidson, and Y. Xing, "Transcriptome sequencing reveals aberrant alternative splicing in huntington's disease," *Hum Mol Genet*, vol. 25, no. 16, pp. 3454–3466, 2016.

[114] H. i. Consortium, "Developmental alterations in huntington's disease neural cells and pharmacological rescue in cells and mice," *Nat Neurosci*, vol. 20, no. 5, pp. 648–660, 2017.

[115] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and . G. P. D. P. Subgroup, "The sequence alignment/map format and samtools," *Bioinformatics*, vol. 25, no. 16, pp. 2078–9, 2009.

[116] P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, and . G. P. A. Group, "The variant call format and vcftools," *Bioinformatics*, vol. 27, no. 15, pp. 2156–8, 2011.

[117] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham, "Plink: a tool set for whole-genome association and population-based linkage analyses," *Am J Hum Genet*, vol. 81, no. 3, pp. 559–75, 2007.

[118] S. D. Turner, "qqman: an r package for visualizing gwas results using q-q and manhattan plots," *bioRxiv*, 2014.

[119] J. C. Barrett, B. Fry, J. Maller, and M. J. Daly, "Haploview: analysis and visualization of ld and haplotype maps," *Bioinformatics*, vol. 21, no. 2, pp. 263–5, 2005.

[120] N. Pourmand, E. Elahi, R. W. Davis, and M. Ronaghi, "Multiplex pyrosequencing," *Nucleic Acids Res*, vol. 30, no. 7, p. e31, 2002.

[121] J. L. Royo, M. Hidalgo, and A. Ruiz, "Pyrosequencing protocol using a universal biotinylated primer for mutation detection and snp genotyping," *Nat Protoc*, vol. 2, no. 7, pp. 1734–9, 2007.

[122] F. Schachter, L. Faure-Delanef, F. GuÈnot, H. Rouger, P. Froguel, L. Lesueur-Ginot, and D. Cohen, "Genetic associations with human longevity at the apoe and ace loci," *Nat Genet*, vol. 6, no. 1, pp. 29–32, 1994.

[123] G. Bu, "Apolipoprotein e and its receptors in alzheimer's disease: pathways, pathogenesis and therapy," *Nature Reviews Neuroscience*, vol. 10, no. 5, pp. 333–344, 2009.

[124] C. R. Scherzer, K. Offe, M. Gearing, H. D. Rees, G. Fang, C. J. Heilman, C. Schaller, H. Bujo, A. I. Levey, and J. J. Lah, "Loss of Apolipoprotein E Receptor LR11 in Alzheimer Disease," *Archives of Neurology*, vol. 61, pp. 1200–1205, 08 2004.

[125] N. Leloup, P. Lössl, D. H. Meijer, M. Brennich, A. J. R. Heck, D. M. E. Thies-Weesie, and B. J. C. Janssen, "Low ph-induced conformational change and dimerization of sortilin triggers endocytosed ligand release," *Nature Communications*, vol. 8, no. 1, p. 1708, 2017.

[126] M. Ostensson, C. MontÈn, J. Bacelis, A. H. Gudjonsdottir, S. Adamovic, J. Ek,

H. Ascher, E. Pollak, H. Arnell, L. Browaldh, D. Agardh, J. Wahlstrˆm, S. Nilsson, and Torinsson-Naluai, "A possible mechanism behind autoimmune disorders discovered by genome-wide linkage and association analysis in celiac disease," *PLoS One*, vol. 8, no. 8, p. e70174, 2013.

[127] K. Foerster and J. Freudenberg, "Hla antigen frequencies in patients with huntington's chorea and their relatives," *J Neurol*, vol. 223, no. 2, pp. 119–25, 1980.

[128] P. Actis, A. Rogers, J. Nivala, B. Vilozny, R. A. Seger, O. Jejelowo, and N. Pourmand, "Reversible thrombin detection by aptamer functionalized sting sensors," *Biosens Bioelectron*, vol. 26, no. 11, pp. 4503–7, 2011.

[129] S. Al Seesi, F. Duan, I. I. Mandoiu, P. K. Srivastava, and A. Kueck, *Genomics-Guided Immunotherapy of Human Epithelial Ovarian Cancer*. ELSEVIER, 2016.

[130] S. Anders, P. T. Pyl, and W. Huber, "Htseq–a python framework to work with high-throughput sequencing data," *Bioinformatics*, vol. 31, no. 2, pp. 166–9, 2015.

[131] S. E. Andrew and M. R. Hayden, "Origins and evolution of huntington disease chromosomes," *Neurodegeneration*, vol. 4, no. 3, pp. 239–44, 1995.

[132] S. Andrew, J. Theilmann, E. Almqvist, A. Norremolle, G. Lucotte, M. Anvret, S. A. Sorensen, J. C. Turpin, and M. R. Hayden, "Dna analysis of distinct populations suggests multiple origins for the mutation causing huntington disease," *Clin Genet*, vol. 43, no. 6, pp. 286–94, 1993.

[133] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P.

Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. the gene ontology consortium," *Nat Genet*, vol. 25, no. 1, pp. 25–9, 2000.

[134] C. T. Ashley and S. T. Warren, "Trinucleotide repeat expansion and human disease," *Annu Rev Genet*, vol. 29, pp. 703–28, 1995.

[135] S. Bassi, T. Tripathi, A. Monziani, F. Di Leva, and M. Biagioli, "Epigenetics of huntington's disease," *Adv Exp Med Biol*, vol. 978, pp. 277–299, 2017.

[136] W. S. Bush and J. H. Moore, "Chapter 11: Genome-wide association studies," *PLoS Comput Biol*, vol. 8, no. 12, p. e1002822, 2012.

[137] D. D. Chaplin, "Overview of the immune response," *J Allergy Clin Immunol*, vol. 125, no. 2 Suppl 2, pp. S3–23, 2010.

[138] G. M. o. H. D. G.-H. Consortium, "Huntington's disease onset is determined by length of uninterrupted cag, not encoded polyglutamine, and is modified by dna maintenance mechanisms," 2019.

[139] M. D. C. Costa, P. Magalhães, L. Guimarães, P. Maciel, J. Sequeiros, and A. Sousa, "The cag repeat at the huntington disease gene in the portuguese population: insights into its dynamics and to the origin of the mutation," *J Hum Genet*, vol. 51, no. 3, pp. 189–195, 2006.

[140] C. J. Cummings and H. Y. Zoghbi, "Trinucleotide repeats: mechanisms and pathophysiology," *Annu Rev Genomics Hum Genet*, vol. 1, pp. 281–328, 2000.

[141] A. R. Di Carluccio, C. F. Triffon, and W. Chen, "Perpetual complexity: predicting human cd8," *Immunol Cell Biol*, vol. 96, no. 4, pp. 358–369, 2018.

[142] M. A. Dillies, A. Rau, J. Aubert, C. Hennequet-Antier, M. Jeanmougin, N. Servant, C. Keime, G. Marot, D. Castel, J. Estelle, G. Guernec, B. Jagla, L. Jouneau, D. Laloë, C. Le Gall, B. Schaëffer, S. Le Crom, M. Guedj, F. Jaffrézic, and F. S. Consortium, "A comprehensive evaluation of normalization methods for illumina high-throughput rna sequencing data analysis," *Brief Bioinform*, vol. 14, no. 6, pp. 671–83, 2013.

[143] R. England and M. Pettersson, "Pyro q-cpg™: quantitative analysis of methylation in multiple cpg sites by pyrosequencing®," *Nature Methods*, vol. 2, 2005.

[144] L. Galluzzi, S. Spranger, E. Fuchs, and A. López-Soto, "Wnt signaling in cancer immunosurveillance," *Trends Cell Biol*, vol. 29, no. 1, pp. 44–65, 2019.

[145] F. Ginhoux, S. Lim, G. Hoeffel, D. Low, and T. Huber, "Origin and differentiation of microglia," *Front Cell Neurosci*, vol. 7, p. 45, 2013.

[146] E. Gjoneska, A. R. Pfenning, H. Mathys, G. Quon, A. Kundaje, L. H. Tsai, and M. Kellis, "Conserved epigenomic signals in mice and humans reveal immune basis of alzheimer's disease," *Nature*, vol. 518, no. 7539, pp. 365–9, 2015.

[147] Y. P. Goldberg, S. E. Andrew, L. A. Clarke, and M. R. Hayden, "A pcr method

for accurate assessment of trinucleotide repeat expansion in huntington disease," *Hum Mol Genet*, vol. 2, no. 6, pp. 635–6, 1993.

[148] Y. P. Goldberg, S. E. Andrew, J. Theilmann, B. Kremer, F. Squitieri, H. Telenius, J. D. Brown, and M. R. Hayden, "Familial predisposition to recurrent mutations causing huntington's disease: genetic risk to sibs of sporadic cases," *J Med Genet*, vol. 30, no. 12, pp. 987–90, 1993.

[149] Y. P. Goldberg, B. Kremer, S. E. Andrew, J. Theilmann, R. K. Graham, F. Squitieri, H. Telenius, S. Adam, A. Sajoo, and E. Starr, "Molecular analysis of new mutations for huntington's disease: intermediate alleles and sex of origin effects," *Nat Genet*, vol. 5, no. 2, pp. 174–9, 1993.

[150] Y. P. Goldberg, C. T. McMurray, J. Zeisler, E. Almqvist, D. Sillence, F. Richards, A. M. Gacy, J. Buchanan, H. Telenius, and M. R. Hayden, "Increased instability of intermediate alleles in families with sporadic huntington disease compared to similar sized intermediate alleles in the general population," *Hum Mol Genet*, vol. 4, no. 10, pp. 1911–8, 1995.

[151] Y. P. Goldberg, C. T. McMurray, J. Zeisler, E. Almqvist, D. Sillence, F. Richards, A. M. Gacy, J. Buchanan, H. Telenius, and M. R. Hayden, "Increased instability of intermediate alleles in families with sporadic huntington disease compared to similar sized intermediate alleles in the general population," *Hum Mol Genet*, vol. 4, no. 10, pp. 1911–8, 1995.

[152] Y. P. Goldberg, J. M. Rommens, S. E. Andrew, G. B. Hutchinson, B. Lin, J. Theilmann, R. Graham, M. L. Glaves, E. Starr, and H. McDonald, "Identification of an alu retrotransposition event in close proximity to a strong candidate gene for huntington's disease," *Nature*, vol. 362, no. 6418, pp. 370–3, 1993.

[153] A. Grover and P. C. Sharma, "Development and use of molecular markers: past and present," *Crit Rev Biotechnol*, vol. 36, no. 2, pp. 290–302, 2016.

[154] R. Guerreiro, A. Wojtas, J. Bras, M. Carrasquillo, E. Rogaeva, E. Majounie, C. Cruchaga, C. Sassi, J. S. Kauwe, S. Younkin, L. Hazrati, J. Collinge, J. Pocock, T. Lashley, J. Williams, J. C. Lambert, P. Amouyel, A. Goate, R. Rademakers, K. Morgan, J. Powell, P. St George-Hyslop, A. Singleton, J. Hardy, and A. G. A. Group, "Trem2 variants in alzheimer's disease," *N Engl J Med*, vol. 368, no. 2, pp. 117–27, 2013.

[155] J. L. Haines, M. A. Hauser, S. Schmidt, W. K. Scott, L. M. Olson, P. Gallins, K. L. Spencer, S. Y. Kwan, M. Noureddine, J. R. Gilbert, N. Schnetz-Boutaud, A. Agarwal, E. A. Postel, and M. A. Pericak-Vance, "Complement factor h variant increases the risk of age-related macular degeneration," *Science*, vol. 308, no. 5720, pp. 419–21, 2005.

[156] T. H. Hamza, C. P. Zabetian, A. Tenesa, A. Laederach, J. Montimurro, D. Yearout, D. M. Kay, K. F. Doheny, J. Paschall, E. Pugh, V. I. Kusel, R. Collura, J. Roberts, A. Griffith, A. Samii, W. K. Scott, J. Nutt, S. A. Factor, and

H. Payami, "Common genetic variation in the hla region is associated with late-onset sporadic parkinson's disease," *Nat Genet*, vol. 42, no. 9, pp. 781–5, 2010.

[157] M. Hosseini, D. Pratas, and A. J. Pinho, "A survey on data compression methods for biological sequences," *Information*, 2016.

[158] J. X. Hu, C. E. Thomas, and S. Brunak, "Network biology concepts in complex disease comorbidities," *Nat Rev Genet*, vol. 17, no. 10, pp. 615–29, 2016.

[159] M. J. Keogh and P. F. Chinnery, "Next generation sequencing for neurological diseases: new hope or new hype?," *Clin Neurol Neurosurg*, vol. 115, no. 7, pp. 948–53, 2013.

[160] K. Kim, S. W. Park, J. H. Kim, S. H. Lee, D. Kim, T. Koo, K. E. Kim, and J. S. Kim, "Genome surgery using cas9 ribonucleoproteins for the treatment of age-related macular degeneration," *Genome Res*, vol. 27, no. 3, pp. 419–426, 2017.

[161] B. Kremer, E. Almqvist, J. Theilmann, N. Spence, H. Telenius, Y. P. Goldberg, and M. R. Hayden, "Sex-dependent mechanisms for expansions and contractions of the cag repeat on affected huntington disease chromosomes," *Am J Hum Genet*, vol. 57, no. 2, pp. 343–50, 1995.

[162] C. W. Law, Y. Chen, W. Shi, and G. K. Smyth, "voom: Precision weights unlock linear model analysis tools for rna-seq read counts," *Genome Biol*, vol. 15, no. 2, p. R29, 2014.

[163] J. T. Leek, "svaseq: removing batch effects and other unwanted noise from sequencing data," *Nucleic Acids Res*, vol. 42, no. 21, 2014.

[164] H. Li, "Tabix: fast retrieval of sequence features from generic tab-delimited files," *Bioinformatics*, vol. 27, no. 5, pp. 718–9, 2011.

[165] G. R. Martho and J. M. Amabis, *Biologia - Das populações - Volume 3Genética, Evolução e Ecologia.* Editora Moderna, 2004.

[166] N. Masuda, J. Goto, N. Murayama, M. Watanabe, I. Kondo, and I. Kanazawa, "Analysis of triplet repeats in the huntingtin gene in japanese families affected with huntington's disease," *J Med Genet*, vol. 32, no. 9, pp. 701–5, 1995.

[167] H. Mi, A. Muruganujan, J. T. Casagrande, and P. D. Thomas, "Large-scale gene function analysis with the panther classification system," *Nat Protoc*, vol. 8, no. 8, pp. 1551–66, 2013.

[168] P. Nyrén, "The history of pyrosequencing," *Methods Mol Biol*, vol. 373, pp. 1–14, 2007.

[169] A. B. Paaby and M. V. Rockman, "The many faces of pleiotropy," *Trends Genet*, vol. 29, no. 2, pp. 66–73, 2013.

[170] H. L. Paulson and K. H. Fischbeck, "Trinucleotide repeats in neurogenetic disorders," *Annu Rev Neurosci*, vol. 19, pp. 79–107, 1996.

[171] M. Richards, "Predictive testing for huntington's disease," *Lancet*, vol. 357, no. 9259, p. 883, 2001.

[172] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edger: a bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, no. 1, pp. 139–40, 2010.

[173] Y. D. Salinas, Z. Wang, and A. T. DeWan, "Statistical analysis of multiple phenotypes in genetic epidemiologic studies: From cross-phenotype associations to pleiotropy," *Am J Epidemiol*, vol. 187, no. 4, pp. 855–863, 2018.

[174] W. Satake, Y. Nakabayashi, I. Mizuta, Y. Hirota, C. Ito, M. Kubo, T. Kawaguchi, T. Tsunoda, M. Watanabe, A. Takeda, H. Tomiyama, K. Nakashima, K. Hasegawa, F. Obata, T. Yoshikawa, H. Kawakami, S. Sakoda, M. Yamamoto, N. Hattori, M. Murata, Y. Nakamura, and T. Toda, "Genome-wide association study identifies common variants at four loci as genetic risk factors for parkinson's disease," *Nat Genet*, vol. 41, no. 12, pp. 1303–7, 2009.

[175] N. Solovieff, C. Cotsapas, P. H. Lee, S. M. Purcell, and J. W. Smoller, "Pleiotropy in complex traits: challenges and strategies," *Nat Rev Genet*, vol. 14, no. 7, pp. 483–95, 2013.

[176] C. Soneson and M. Delorenzi, "A comparison of methods for differential expression analysis of rna-seq data," *BMC Bioinformatics*, vol. 14, p. 91, 2013.

[177] F. Squitieri, S. E. Andrew, Y. P. Goldberg, B. Kremer, N. Spence, J. Zeisler, K. Nichol, J. Theilmann, J. Greenberg, and J. Goto, "Dna haplotype analysis of huntington disease reveals clues to the origins and mechanisms of cag expansion

and reasons for geographic variations of prevalence," *Hum Mol Genet*, vol. 3, no. 12, pp. 2103–14, 1994.

[178] N. Z. Steele, J. S. Carr, L. W. Bonham, E. G. Geier, V. Damotte, Z. A. Miller, R. S. Desikan, K. L. Boehme, S. Mukherjee, P. K. Crane, J. S. Kauwe, J. H. Kramer, B. L. Miller, G. Coppola, J. A. Hollenbach, Y. Huang, and J. S. Yokoyama, "Fine-mapping of the human leukocyte antigen locus as a risk factor for alzheimer disease: A case-control study," *PLoS Med*, vol. 14, no. 3, p. e1002272, 2017.

[179] X. Sui, R. Chen, Z. Wang, Z. Huang, N. Kong, M. Zhang, W. Han, F. Lou, J. Yang, Q. Zhang, X. Wang, C. He, and H. Pan, "Autophagy and chemotherapy resistance: a promising therapeutic target for cancer treatment," *Cell Death Dis*, vol. 4, p. e838, 2013.

[180] P. D. Thomas, M. J. Campbell, A. Kejariwal, H. Mi, B. Karlak, R. Daverman, K. Diemer, A. Muruganujan, and A. Narechania, "Panther: a library of protein families and subfamilies indexed by function," *Genome Res*, vol. 13, no. 9, pp. 2129–41, 2003.

[181] J. van Loosdregt and P. J. Coffer, "The role of wnt signaling in mature t cells: T cell factor is coming home," *J Immunol*, vol. 201, no. 8, pp. 2193–2200, 2018.

[182] J. S. Yokoyama, Y. Wang, A. J. Schork, W. K. Thompson, C. M. Karch, C. Cruchaga, L. K. McEvoy, A. Witoelar, C. H. Chen, D. Holland, J. B. Brewer, A. Franke, W. P. Dillon, D. M. Wilson, P. Mukherjee, C. P. Hess, Z. Miller, L. W.

Bonham, J. Shen, G. D. Rabinovici, H. J. Rosen, B. L. Miller, B. T. Hyman, G. D. Schellenberg, T. H. Karlsen, O. A. Andreassen, A. M. Dale, R. S. Desikan, and A. D. N. Initiative, "Association between genetic traits for immune-mediated diseases and alzheimer disease," *JAMA Neurol*, vol. 73, no. 6, pp. 691–7, 2016.

[183] L. Zhao, S. H. Patel, J. Pei, and K. Zhang, "Antagonizing wnt pathway in diabetic retinopathy," *Diabetes*, vol. 62, no. 12, pp. 3993–5, 2013.

[184] F. Zou, X. Wang, X. Han, G. Rothschild, S. G. Zheng, U. Basu, and J. Sun, "Expression and function of tetraspanins and their interacting partners in b cells," *Front Immunol*, vol. 9, p. 1606, 2018.

[185] C. Zühlke, O. Riess, B. Bockel, H. Lange, and U. Thies, "Mitotic stability and meiotic variability of the (cag)n repeat in the huntington disease gene," *Hum Mol Genet*, vol. 2, no. 12, pp. 2063–7, 1993.

[186] K. J. Arrow, L. Hurwicz, and H. Uzawa, "Constraint qualifications in maximization problems," *Naval Research Logistics Quarterly*, vol. 8, pp. 175–191, 1961.

[187] R. Nambron, E. Silajdžić, E. Kalliolia, C. Ottolenghi, P. Hindmarsh, and N. R. Hill, "A metabolic study of huntington's disease," *PLoS One*, vol. 11, 2016.

[188] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for rna-seq data with deseq2," *Genome Biol*, vol. 15, p. 550, 2014.

[189] C. Reitz, S. Tokuhiro, L. N. Clark, C. Conrad, J.-P. P. Vonsattel, and L.-N. N.

Hazrati, "Sorcs1 alters amyloid precursor protein processing and variants may increase alzheimer's disease risk," *Ann. Neurol.*, vol. 69, pp. 47–64, 2011.

[190] J. C. Hou and J. E. Pessin, "Ins (endocytosis) and outs (exocytosis) of glut4 trafficking," *Curr. Opin. Cell. Biol.*, vol. 19, pp. 466–473, 2007.

[191] L. A. Farrer, "Diabetes mellitus in huntington disease," *Clin. Genet.*, vol. 27, pp. 62–67, 1985.

[192] P. W. Sutter and E. A. Sutter, "Dispensing and and surface-induced crystallization of zeptolitre liquid metal-alloy drops," *Nat. Mater.*, vol. 6, pp. 363–366, 2007.

[193] F. O. Laforge, J. Carpino, S. A. Rotenberg, and M. V. Mirkin, "Electrochemical attosyringe," *Proc. Natl. Acad. Sci USA*, vol. 104, pp. 11895–11900, 2007.

[194] H. Matsuoka, T. Komazaki, Y. Mukai, M. Shibusawa, H. Akane, A. Chaki, N. Uetake, and M. Saito, "High throughput easy microinjection with a single-cell manipulation supporting robot," *J. Biotechnol*, vol. 116, pp. 185–19, 2005.

[195] B. Treutlein, D. G. Brownfield, A. R. Wu, N. F. Neff, G. L. Mantalas, F. H. Espinoza, T. J. Desai, M. A. Krasnow, and S. R. Quake, "Reconstructing lineage hierarchies of the distal lung epithelium using single-cell rna-seq," *Nature*, vol. 509, pp. 371–375, 2014.

[196] A. M. Streets, X. Zhang, C. Cao, Y. Pang, X. Wu, L. Xiong, L. Yang, Y. Fu, L. Zhao, F. Tang, and Y. Huang, "Microfluidic single-cell whole-transcriptome sequencing," in *Proc. Natl. Acad. Sci. USA 111*, pp. 7048–7053, 2014.

[197] H. Uehara, T. Osada, and A. Ikai, "Quantitative measurement of mrna at different loci within an individual living cell," *Ultramicroscopy*, vol. 100, pp. 197–201, 2004.

[198] M.-C. W. Lee, F. J. Lopez-Diaz, S. Y. Khan, M. A. Tariq, Y. Dayn, C. J. Vaske, A. J. Radenbaugh, H. J. Kim, B. M. Emerson, and N. Pourmand, "Single-cell analyses of transcriptional heterogeneity during drug tolerance transition in cancer cells by rna sequencing," in *Proc. Natl. Acad. Sci. USA 111*, (E4726–E47), 2014.

[199] P. Actis, A. C. Mak, and N. Pourmand, "Functionalized nanopipettes: Toward label-free single cell biosensors," *Bioanal. Rev*, vol. 1, pp. 177–185, 2010.

[200] M. Karhanek, J. T. Kemp, N. Pourmand, R. W. Davis, and C. D. Webb, "Single dna molecule detection using nanopipettes and nanoparticles.," *Nano. Lett.*, vol. 5, pp. 403–407, 2005.

[201] P. Actis, M. M. Maalouf, H. J. Kim, A. Lohith, B. Vilozny, R. A. Seger, and N. Pourmand, "Compartmental genomics in living cells revealed by single-cell nanobiopsy," *ACS Nano*, vol. 8, pp. 546–553, 2013.

[202] B. Vilozny, P. Actis, R. A. Seger, and N. Pourmand, "Dynamic control of nanoprecipitation in a nanopipette," *ACS Nano*, vol. 5, pp. 3191–3197, 2011.

[203] L. Ying, "Applications of nanopipettes in bionanotechnology;," *Portland Press Limited: London, UK*, vol. 2009.

[204] C. A. Morris, A. K. Friedman, and L. A. Baker, "Applications of nanopipettes in the analytical sciences," *Analyst*, vol. 135, pp. 2190–2202, 2010.

[205] T. Takami, B. H. Park, and T. Kawai, "Nanopipette exploring nanoworld," *Nano Converg.*, vol. 1, 2014.

[206] E. Neher and B. Sakmann, "Single-channel currents recorded from membrane of denervated frog muscle fibres," *Nature*, vol. 260, pp. 799–802, 1976.

[207] M. Mirkin, S. Amemiya, and M. Mirkin, "Nanoelectrodes and liquid/liquid nanointerfaces," in *Nanoelectrochemistry*, pp. 539–572, CRC Press Boca Raton, FL, USA, 2015.

[208] C. Trapnell, D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N. J. Lennon, K. J. Livak, T. S. Mikkelsen, and J. L. Rinn, "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells," *Nat. Biotechnol.*, vol. 32, pp. 381–386, 2014.

[209] M. H. Spitzer and G. Nolan, "Mass cytometry: Single cells and many features," *Cell*, vol. 165, pp. 780–791, 2016.

[210] D. J. Stephens and R. Pepperkok, "The many ways to cross the plasma membrane," in *Proc. Natl. Acad. Sci. USA 98*, pp. 4295–4298, 2001.

[211] A. Meister, M. Gabi, P. Behr, P. Studer, J. Vörös, P. Niedermann, J. Bitterli, J. Polesel-Maris, M. Liley, H. Heinzelmann, *et al.*, "Fluidfm: Combining atomic force microscopy and nanofluidics in a universal liquid delivery system for single cell applications and beyond," *Nano. Lett.*, vol. 9, pp. 2501–2507, 2009.

[212] K. T. Rodolfa, A. Bruckbauer, D. Zhou, Y. E. Korchev, and D. Klenerman, "Two-component graded deposition of biomolecules with a double-barreled nanopipette," *Angew. Chem Int. Ed*, vol. 44, pp. 6854–6859, 2005.

[213] Y.-C. Wu, T.-H. Wu, D. L. Clemens, B.-Y. Lee, X. Wen, M. A. Horwitz, M. A. Teitell, and P.-Y. Chiou, "Massively parallel delivery of large cargo into mammalian cells with light pulses," *Nat. Methods*, vol. 12, pp. 439–444, 2015.

[214] T.-H. Wu, E. Sagullo, D. Case, X. Zheng, Y. Li, J. S. Hong, T. TeSlaa, A. N. Patananan, J. M. McCaffery, K. Niazi, *et al.*, "Mitochondrial transfer by photothermal nanoblade restores metabolite profile in mammalian cells," *Cell Metab*, vol. 23, pp. 921–929, 2016.

[215] X. Li, Y. Tao, D.-H. Lee, H. K. Wickramasinghe, and A. P. Lee, "In situ mrna isolation from a microfluidic single-cell array using an external afm nanoprobe," *Lab Chip*, vol. 17, pp. 1635–1644, 2017.

[216] Y.-C. Chen, H. W. Baac, K.-T. Lee, S. Fouladdel, K. Teichert, J. G. Ok, Y.-H. Cheng, P. N. Ingram, A. J. Hart, and E. Azizi, "Selective photomechanical detachment and retrieval of divided sister cells from enclosed microfluidics for downstream analyses," *ACS Nano*, vol. 11, pp. 4660–4668, 2017.

[217] R. Yan, J.-H. Park, Y. Choi, C.-J. Heo, S.-M. Yang, L. P. Lee, and P. Yang, "Nanowire-based single-cell endoscopy," *Nat. Nanotechnol.*, vol. 7, pp. 191–196, 2012.

[218] R. Singhal, Z. Orynbayeva, R. V. K. Sundaram, J. J. Niu, S. Bhattacharyya, E. A. Vitol, M. G. Schrlau, E. S. Papazoglou, G. Friedman, and Y. Gogotsi, "Multifunctional carbon-nanotube cellular endoscopes," *Nat. Nanotechnol.*, vol. 6, pp. 57–64, 2011.

[219] J. König, K. Zarnack, N. M. Luscombe, and J. Ule, "Protein-rna interactions: New genomic technologies and perspectives," *Nat. Rev. Genet.*, vol. 13, pp. 77–83, 2012.

[220] D. M. Chudakov, S. Lukyanov, and K. A. Lukyanov, "Fluorescent proteins as a toolkit for in vivo imaging," *Trends Biotechnol*, vol. 23, pp. 605–613, 2005.

[221] E. B. Voura, J. K. Jaiswal, H. Mattoussi, and S. M. Simon, "Tracking metastatic tumor cell extravasation with quantum dot nanocrystals and fluorescence emission-scanning microscopy," *Nat. Med.*, vol. 10, pp. 993–998, 2004.

[222] J. P. Junker, B. Spanjaard, J. Peterson-Maduro, A. Alemany, B. Hu, M. Florescu, and A. van Oudenaarden, "Massively parallel clonal analysis using crispr/cas induced genetic scars.," *bioRxiv*, vol. 9, 2017.

[223] K. Kretzschmar and F. M. Watt, "Lineage tracing," *Cell*, vol. 148, pp. 33–45, 2012.

[224] D. Nawarathna, R. Chang, E. Nelson, and H. K. Wickramasinghe, "Targeted mrna profiling of transfected breast cancer gene in a living cell," *Anal. Biochem.*, vol. 408, pp. 342–344, 2011.

[225] H. Uehara, Y. Kunitomi, A. Ikai, and T. Osada, "mrna detection of individual cells with the single cell nanoprobe method compared with in situ hybridization," *J. Nanobiotechnol.*, vol. 5, p. 7, 2007.

[226] I. E. Clark, M. W. Dodson, C. Jiang, J. H. Cao, J. R. Huh, J. H. Seol, S. J. Yoo, B. A. Hay, and M. Guo, "Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin," *Nature*, vol. 441, pp. 1162–1166, 2006.

[227] A. Ståhlberg, C. Thomsen, D. Ruff, and P. Åman, "Quantitative pcr analysis of dna, rnas, and proteins in the same single cell," *Clin. Chem.*, vol. 58, pp. 1682–1691, 2012.

[228] L. Jiang, F. Schlesinger, C. A. Davis, Y. Zhang, R. Li, M. Salit, T. R. Gingeras, and B. Oliver, "Synthetic spike-in standards for rna-seq experiments," *Genome Res.*, vol. 21, pp. 1543–1551, 2011.

[229] C. Klijn, S. Durinck, E. W. Stawiski, P. M. Haverty, Z. Jiang, H. Liu, J. Degenhardt, O. Mayba, F. Gnad, J. Liu, *et al.*, "A comprehensive transcriptional portrait of human cancer cell lines," *Nat. Biotechnol.*, vol. 33, pp. 306–312, 2015.

[230] P. Ramdas, M. Rajihuzzaman, S. D. Veerasenan, K. R. Selvaduray, K. Nesaretnam, and A. K. Radhakrishnan, "Tocotrienol-treated mcf-7 human breast cancer cells show down-regulation of api5 and up-regulation of mig6 genes," *Cancer Genom. Proteom.*, vol. 8, pp. 19–31, 2011.

[231] M. B. Lyng, A.-V. Lænkholm, N. Pallisgaard, and H. J. Ditzel, "Identification

of genes for normalization of real-time rt-pcr data in breast carcinomas," *BMC Cancer*, vol. 8, p. 20, 2008.

[232] A. Shadeo and W. L. Lam, "Comprehensive copy number profiles of breast cancer cell model genomes," *Breast Cancer Res.*, vol. 8, 2006.

[233] R. M. Linger, A. K. Keating, H. S. Earp, and D. K. Graham, "Tam receptor tyrosine kinases: Biologic functions signaling, and potential therapeutic targeting in human cancer," *Adv. Cancer Res.*, vol. 100, pp. 35–83, 2008.

[234] X. Zhang, R. Schulz, S. Edmunds, E. Krüger, E. Markert, J. Gaedcke, E. Cormet-Boyaka, M. Ghadimi, T. Beissbarth, A. J. Levine, *et al.*, "Microrna-101 suppresses tumor cell proliferation by acting as an endogenous proteasome inhibitor via targeting the proteasome assembly factor pomp," *Mol. Cell*, vol. 59, pp. 243–257, 2015.

[235] J. Gorelik, Y. Gu, H. A. Spohr, A. I. Shevchuk, M. J. Lab, S. E. Harding, C. R. W. Edwards, M. Whitaker, G. W. J. Moss, D. C. H. Benton, *et al.*, "Ion channels in small cells and subcellular structures can be studied with a smart patch-clamp system," *Biophys. J.*, vol. 83, pp. 3296–3303, 2002.

[236] K. Svensson, H. Olin, and E. Olsson, "Nanopipettes for metal transport," *Phys. Rev. Lett.*, vol. 93, p. 145901, 2004.

[237] L. Ying, A. Bruckbauer, D. Zhou, J. Gorelik, A. Shevchuk, Y. Korchev, and D. Klenerman, "The scanned nanopipette: A new tool for high resolution bioimag-

ing and controlled deposition of biomolecules," *Phys. Chem. Chem. Phys.*, vol. 7, pp. 2859–2866, 2005.

[238] S. Umehara, N. Pourmand, C. D. Webb, R. W. Davis, K. Yasuda, and M. Karhanek, "Current rectification with poly-l-lysine-coated quartz nanopipettes," *Nano Lett.*, vol. 6, pp. 2486–2492, 2006.

[239] K. T. Rodolfa, A. Bruckbauer, D. Zhou, A. I. Schevchuk, Y. E. Korchev, and D. Klenerman, "Nanoscale pipetting for controlled chemistry in small arrayed water droplets using a double-barrel pipet," *Nano Lett.*, vol. 6, pp. 252–257, 2006.

[240] F. Iwata, S. Nagami, Y. Sumiya, and A. Sasaki, "Nanometre-scale deposition of colloidal au particles using electrophoresis in a nanopipette probe," *Nanotechnology*, vol. 18, no. 10530, p. 1, 2007.

[241] A. Bruckbauer, P. James, D. Zhou, J. W. Yoon, D. Excell, Y. Korchev, R. Jones, and D. Klenerman, "Nanopipette delivery of individual molecules to cellular compartments for single-molecule fluorescence tracking," *Biophys. J.*, vol. 93, pp. 3120–3131, 2007.

[242] C. K. Byun, X. Wang, Q. Pu, and S. E.-B. N. Liu, "Electroosmosis-based nanopipettor," *Anal. Chem.*, vol. 79, pp. 3862–3866, 2007.

[243] M. Karhanek, C. D. Webb, S. Umehara, and N. Pourmand, "Functionalized nanopipette biosensor," *U.S. Patent Application No. 20100072080*, vol. 2080, p. 2010007, March 2010.

[244] A. Rothery, J. Gorelik, A. Bruckbauer, W. Yu, Y. Korchev, and D. Klenerman, "A novel light source for sicm-snom of living cells," *J. Microsc.*, vol. 209, pp. 94–101, 2003.

[245] J. Gorelik, Y. Zhang, A. I. Shevchuk, G. I. Frolenkov, D. Sánchez, I. Vodyanoy, C. R. Edwards, D. Klenerman, and Y. E. Korchev, "The use of scanning ion conductance microscopy to image a6 cells," *Mol. Cell. Endocrinol.*, vol. 217, pp. 101–108, 2004.

[246] U. Windhorst and H. Johansson, "Modern techniques in neuroscience research," *Springer Science Business Media: Berlin, Germany*, 2012.

[247] L. Ying, S. S. White, A. Bruckbauer, L. Meadows, Y. E. Korchev, and D. Klenerman, "Frequency and voltage dependence of the dielectrophoretic trapping of short lengths of dna and dctp in a nanopipette," *Biophys. J.*, vol. 86, pp. 1018–1027, 2004.

[248] L. Ying, A. Bruckbauer, A. M. Rothery, Y. E. Korchev, and D. Klenerman, "Programmable delivery of dna through a nanopipet," *Anal. Chem.*, vol. 74, pp. 1380–1385, 2002.

[249] R. W. Clarke, S. S. White, D. Zhou, L. Ying, and D. Klenerman, "Trapping of proteins under physiological conditions in a nanopipette," *Angew. Chem. Int. Ed.*, vol. 44, pp. 3747–3750, 2005.

[250] R.-J. Yu, Y.-L. Ying, Y.-X. Hu, R. Gao, and Y.-T. Long, "Label-free monitoring

of single molecule immunoreaction with a nanopipette," *Anal. Chem.*, vol. 89, pp. 8203–8206, 2017.

[251] J. P. Chambers, B. P. Arulanandam, L. L. Matta, A. Weis, and J. J. Valdes, "Biosensor recognition elements;," *Department of Biology, University of Texas at San Antonio: San Antonio, TX, USA*, 2008.

[252] V. Thiviyanathan and D. G. Gorenstein, "Aptamers and the next generation of diagnostic reagents," *Proteom. Clin. Appl.*, vol. 6, pp. 563–573, 2012.

[253] A. Z. Wang and O. C. Farokhzad, "Current progress of aptamer-based molecular imaging," *J. Nucl. Med*, vol. 55, pp. 353–356, 2014.

[254] P. Actis, A. Rogers, J. Nivala, B. Vilozny, R. A. Seger, O. Jejelowo, and N. Pourmand, "Reversible thrombin detection by aptamer functionalized sting sensors," *Biosens. Bioelectron.*, vol. 26, pp. 4503–4507, 2011.

[255] V. Heiden, M. G. Cantley, L. C. Thompson, and C. B. U. the Warburg Effect, "The metabolic requirements of cell proliferation," *Science*, vol. 324, pp. 1029–1033, 2009.

[256] A. Annibaldi and C. Widmann, "Glucose metabolism in cancer cells," *Curr. Opin. Clin. Nutr. Metab. Care*, vol. 13, pp. 466–470, 2010.

[257] T. Wu, C. T. Sempos, J. L. Freudenheim, P. Muti, and E. Smit, "Serum iron, copper and zinc concentrations and risk of cancer mortality in us adults," *Ann. Epidemiol.*, vol. 14, pp. 195–201, 2004.

[258] H. Wiseman and B. Halliwell, "Damage to dna by reactive oxygen and nitrogen species: Role in inflammatory disease and progression to cancer," *Biochem. J.*, 1996.

[259] R. A. Gatenby and R. J. Gillies, "Why do cancers have high aerobic glycolysis," *Nat. Rev. Cancer*, vol. 4, pp. 891–899, 2004.

[260] N. Azad, Y. Rojanasakul, and V. Vallyathan, "Inflammation and and lung cancer: Roles of reactive oxygen/nitrogen species," *J. Toxicol. Environ*, vol. 11, pp. 1–15, 2008.

[261] C. C. Winterbourn, "The challenges of using fluorescent probes to detect and quantify specific reactive oxygen species in living cells," *Biochim. Biophys. Acta (BBA) Gen. Subj.*, vol. 1840, pp. 730–738, 2014.

[262] L. J. Steinbock, O. Otto, C. Chimerel, J. Gornall, and U. F. Keyser, "Detecting dna folding with nanocapillaries," *Nano Lett.*, vol. 10, pp. 2493–2497, 2010.

[263] A. V. A. P. Q.-T. F. A. J. J. B. Gong, X.;Patil, "Label-free in-flow detection of single dna molecules using glass nanopipettes.," *Anal. Chem.*, vol. 86, pp. 835–841, 2014.

[264] B. Vilozny, P. Actis, R. A. Seger, Q. Vallmajo-Martin, and N. Pourmand, "Reversible cation response with a protein-modified nanopipette," *Anal. Chem.*, vol. 83, pp. 6121–6126, 2011.

[265] N. Sa, Y. Fu, and L. A. Baker, "Reversible cobalt ion binding to imidazole-modified nanopipettes," *Anal. Chem.*, vol. 82, pp. 9963–9966, 2010.

[266] Y. Cao, M. Hjort, H. Chen, F. Birey, S. Leal-Ortiz, C. Han, J. Santiago, S. Pas ca, J. Wu, and N. Melosh, "Nondestructive nanostraw intracellular sampling for longitudinal cell monitoring.," *Proc. Natl. Acad. Sci. USA*, vol. 114, p. E1866–E1874, 2017.

[267] O. Kanisicak, H. Khalil, M. J. Ivey, J. Karch, B. D. Maliken, R. N. Correll, M. J. Brody, S.-C. J. Lin, B. J. Aronow, M. D. Tallquist, *et al.*, "Genetic lineage tracing defines myofibroblast origin and function in the injured heart," *Nat. Commun.*, vol. 7, p. 12260, 2016.

[268] R. A. Beckman, G. S. Schemmann, and C.-H. Yeang, "Impact of genetic dynamics and single-cell heterogeneity on development of nonstandard personalized medicine strategies for cancer," in *Proc. Natl. Acad. Sci. USA 109*, pp. 14586–14591, 2012.

[269] G. Housman, S. Byler, S. Heerboth, K. Lapinska, M. Longacre, N. Snyder, and S. Sarkar, "Drug resistance in cancer: An overview," *Cancers*, vol. 6, pp. 1769–1792, 2014.

[270] K. O. Alfarouk, C.-M. Stock, S. Taylor, M. Walsh, A. K. Muddathir, D. Verduzco, A. H. Bashir, O. Y. Mohammed, G. O. Elhassan, S. Harguindey, *et al.*, "Resistance

to cancer chemotherapy: Failure in drug response from adme to p-gp," *Cancer Cell Int.*, vol. 15, p. 71, 2015.

[271] N. L. Komarova and D. Wodarz, "Drug resistance in cancer: Principles of emergence and prevention," in *Proc. Natl. Acad. Sci. USA 102*, pp. 9714–9719, 2005.

[272] X. Sui, R. Chen, Z. Wang, Z. Huang, N. Kong, M. Zhang, W. Han, F. Lou, J. Yang, Q. Zhang, *et al.*, "Autophagy and chemotherapy resistance: A promising therapeutic target for cancer treatment," *Cell Death Dis.*, vol. 4, 2013.

[273] P. Borst, "Cancer drug pan-resistance: Pumps, cancer stem cells, quiescence, epithelial to mesenchymal transition, blocked cell death pathways, persisters or what?," *Open Biol.*, vol. 2, no. 12006, p. 6, 2012.

[274] T. E. Creighton, "Protein structure: A practical approach," *Oxford University Press: Oxford, NY, USA*, 1997.

[275] O. H. Lowry, N. J. Rosebrough, A. L. Farr, and R. J. Randall, "Protein measurement with the folin phenol reagent," *J. Biol. Chem.*, vol. 193, pp. 265–275, 1951.

[276] M. M. Bradford, "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding," *Anal. Biochem.*, vol. 72, pp. 248–254, 1976.

[277] J. Renart, J. Reiser, and G. R. Stark, "Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: A method for study-

ing antibody specificity and antigen structure," in *Proc. Natl. Acad. Sci. USA 76*, pp. 3116–3120, 1979.

[278] A. J. Hughes, D. P. Spelke, Z. Xu, C.-C. Kang, D. V. Schaffer, and A. E. Herr, "Single-cell western blotting," *Nat. Methods*, vol. 11, pp. 749–755, 2014.

[279] W. Wei, Y. S. Shin, C. Ma, J. Wang, M. Elitas, R. Fan, and J. R. Heath, "Microchip platforms for multiplex single-cell functional proteomics with applications to immunology and cancer research," *Genome Med*, vol. 5, p. 75, 2013.

[280] S. Umehara, M. Karhanek, R. W. Davis, and N. Pourmand, "Label-free biosensing with functionalized nanopipette probes," in *Proc. Natl. Acad. Sci. USA 106*, pp. 4611–4616, 2009.

[281] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, *et al.*, "The sequence of the human genome," *Science*, vol. 291, pp. 1304–1351, 2001.

[282] J. W. Young, J. C. W. Locke, A. Altinok, N. Rosenfeld, T. Bacarian, P. S. Swain, E. Mjolsness, and M. B. Elowitz, "Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy," *Nat. Protoc.*, vol. 7, pp. 80–88, 2012.

[283] I. Glauche, M. Herberg, and I. Roeder, "Nanog variability and and pluripotency regulation of embryonic stem cells-insights from a mathematical model analysis," *PLoS ONE*, vol. 5, 2010.

[284] S. E. C. Dale and P. R. Unwin, "Polarised liquid/liquid micro-interfaces move during charge transfer," *Electrochem. Commun.*, vol. 10, pp. 723–726, 2008.

[285] E. N. Tóth, A. Lohith, M. Mondal, J. Guo, A. Fukamizu, and N. Pourmand, "Single-cell nanobiopsy reveals compartmentalization of mrna in neuronal cells," *J. Biol. Chem.*, vol. 293, pp. 4940–4951, 2018.

[286] Y. Nashimoto, Y. Takahashi, Y. Zhou, H. Ito, H. Ida, K. Ino, T. Matsue, and H. Shiku, "Evaluation of mrna localization using double barrel scanning ion conductance microscopy," *ACS Nano*, vol. 10, pp. 6915–6922, 2016.

[287] O. Guillaume-Gentil, R. V. Grindberg, R. Kooger, L. Dorwling-Carter, V. Martinez, D. Ossola, M. Pilhofer, T. Zambelli, and J. A. Vorholt, "Tunable single-cell extraction for molecular analyses," *Cell*, vol. 166, pp. 506–516, 2016.

[288] X. Luo and L. Li, "Metabolomics of small numbers of cells: Metabolomic profiling of 100, 1000, and 10000 human breast cancer cells," *Anal. Chem.*, vol. 89, pp. 11664–11671, 2017.

[289] O. Guillaume-Gentil, T. Rey, P. Kiefer, A. J. Ibáñez, R. Steinhoff, R. Brönnimann, L. Dorwling-Carter, T. Zambelli, R. Zenobi, and J. A. Vorholt, "Single-cell mass spectrometry of metabolites extracted from live cells by fluidic force microscopy," *Anal. Chem.*, vol. 89, pp. 5017–5023, 2017.

[290]

[291] F. Sch‰chter, L. Faure-Delanef, F. GuÈnot, H. Rouger, P. Froguel, L. Lesueur-Ginot, and D. Cohen, "Genetic associations with human longevity at the apoe and ace loci," *Nat Genet*, vol. 6, no. 1, pp. 29–32, 1994.

[292] M. ÷stensson, C. MontÈn, J. Bacelis, A. H. Gudjonsdottir, S. Adamovic, J. Ek, H. Ascher, E. Pollak, H. Arnell, L. Browaldh, D. Agardh, J. Wahlstrˆm, S. Nilsson, and Torinsson-Naluai, "A possible mechanism behind autoimmune disorders discovered by genome-wide linkage and association analysis in celiac disease," *PLoS One*, vol. 8, no. 8, p. e70174, 2013.

[293] H. i. Consortium, "Developmental alterations in huntington's disease neural cells and pharmacological rescue in cells and mice," *Nat Neurosci*, vol. 20, no. 5, pp. 648–660, 2017.

[294] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo, "The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data," *Genome Res*, vol. 20, no. 9, pp. 1297–303, 2010.

Appendix A

Review Article: Nanopipettes As Monitoring Probes For The Single Living Cell: State Of The Art And Future Directions In Molecular Biology

Examining the behavior of a single cell within its natural environment is valuable for understanding both the biological processes that control the function of cells and how injury and disease lead to pathological change of their function. Single-cell analysis can reveal information regarding the causes of genetic changes, and can contribute to studies on the molecular basis of cell transformation and proliferation. In contrast, whole tissue biopsies can only yield information on a statistical average of several processes occurring in a population of different cells. Electrowetting within a

nanopipette provides a nanobiopsy platform for the extraction of cellular material from single living cells. Additionally, functionalized nanopipette sensing probes can differentiate analytes based on their size, shape or charge density, making the technology uniquely suited to sensing changes in single-cell dynamics. In this chapter, we highlight the potential of nanopipette technology as a non-destructive analytical tool to monitor single living cells, with particular attention to integration into applications in molecular biology.

A.1 Introduction

Nanopipettes are of scientific interest due to their application potential in several arenas, from biomedical diagnostics to cellular biology. Nanopipettes are characterized by the submicron to nanoscale size of the pore opening at the tip, which serves as a suitable surface to fabricate functional tools for delivery to and/or aspiration from a single living cell, or for probing the cell's contents. The hollow structure enables the dispensation of fluid from one region to the next, with their cavity acting as passage [192]. In view of the fact that many biologically significant molecules, such as DNA and proteins, are not able to spontaneously cross the cell membrane [193], the use of a non-destructive single cell manipulation platform such as nanopipettes to study single-cell dynamics is rapidly increasing. Other analysis techniques that require dissociation of tissue from its natural environment lead to the loss of spatial information on individual cells. Previous efforts at single cell manipulation include microinjection

to introduce molecules into the cytoplasm of single cells [194]; microfluidic technologies [196], scanning probe and atomic force microscopy [197] to extract various biomolecules from the cell cytosol. Nanopipettes offer significant advantages over these techniques in that they target a specific single cell and the particular parts of the cell, including the nucleus, and the ability to inject the cargo precisely. The fundamental understanding of the molecular biology of single living cells in heterogeneous cell populations is of the utmost importance in assessing changes in cellular functions in tissues. Whole tissue biopsies can provide information on many events that are occurring in different cells, but difficulties not always suitable for drawing conclusions regarding the progression of some diseases. For example, malignant tumors are heterogeneous in most cases and can include cells at different stages of transformation [198]. Because they provide a tool that both can inject molecules into a cell and probe for the presence of biomarker molecules, nanopipettes are useful in correlating the cellular mechanism of one disease with another, as was recently demonstrated for Huntington's and intracellular glucose levels [28]. Thus, the use of multi-functional nanopipettes in single cell interrogation is beneficial in understanding the mechanism and pathways that link two related conditions, aiding in the development of drug therapies, and at the same time contributing to diagnostics for at-risk individuals. Tools such as nanopipettes, which are easy to adapt to several fields by modifying the nanopipette with different functionalities, can find application in many scientific disciplines [199, 200, 201, 202, 203]. Pipettes have been employed to transfer specified volumes of liquids in science and medicine for centuries [204]. The use of glass micropipette as an intracellular microelectrode was shown as

early as 1902 [205]. Later, the increasing need for precise manipulation of small volumes in molecular biology resulted in the production of micropipettes with the ability to dispense volumes in the µL to mL range. Pipettes were used in the patch-clamp method in 1976 by Neher and Sakmann for detection of voltages and current from ion-channels [206]. Most recently, with the advances in electrophysiology and manufacturing at the nanoscale, nanopipettes emerged as useful tools for both in controlling and depositing small volumes, and in analytical sciences. Previous publications have summarized the production and characterization of different types of nanopipettes [207]. In this chapteer, we focus on the different areas of application of nanopipettes in molecular biology, which include their use as: (1) surgical tools to inject or aspirate molecules from single living cells; (2) functional probes to monitor the presence of biologically relevant molecules in single cells.

A.2 Use of Nanopipettes as Surgical Tools

A.2.1 Nanoinjections by Single-Cell Surgery

Recently, information illuminating the behavior of single cells has received a great deal of attention [208, 209]. To assess the response of a single cell, it is necessary to have an instrument capable of rapidly analyzing and manipulating individual cells in an automated way, while avoiding any damage that could affect these cells' viability. Conventional methods of cell injection employ micropipettes [210] that deliver a large volume of substance that is incompatible with the size of typical cells. Other methods,

such as atomic force microscopy (AFM), hollow cantilevers [211] were constructed, but are also limited by lack of control of injection volumes. Electrochemical autosyringes that deliver the cargo by applying voltage across the liquid/liquid interface [193] and double-barrel glass nanopipettes capable of controlled deposition of biomolecules onto functionalized surfaces [212] showed potential for injections through cell membranes. Previous studies examined the injections of and aspiration from the cells based on microfluidic devices, light pulses, and photothermal nanoblades [213, 214, 215, 216]. The main advantage of our nanopipette/nanosensor over these methods is the electrical system, which has a built-in feedback mechanism using homemade software that allows the user to find the cell and penetrate in an automated manner. Our group has presented the development of a fully electrical system that makes it possible to inject a controlled amount of material (50 fL) into a single cell [201]. We demonstrated the deployment of the system with injections of fluorescent dyes into adherent mammalian cells [9]. A Scanning Ion Conductance Microscope (SICM) was integrated in the platform so we could differentiate a single cell from the population of cells. The system can detect the target cell surface and enables the delivery of molecules into individual cells by employing voltage pulses. Unique advantages of our system include simultaneous cell surface detection and control over the volume of cargo delivery, which were not included in previous nanoscale injection systems [217, 218]. A nanopipette is used in our cell injection system both for feedback-controlled nano-positioning and for delivery. The system is designed to work by following five steps: approach, feedback, penetration, injection, and retraction, as depicted in Figure A.1.

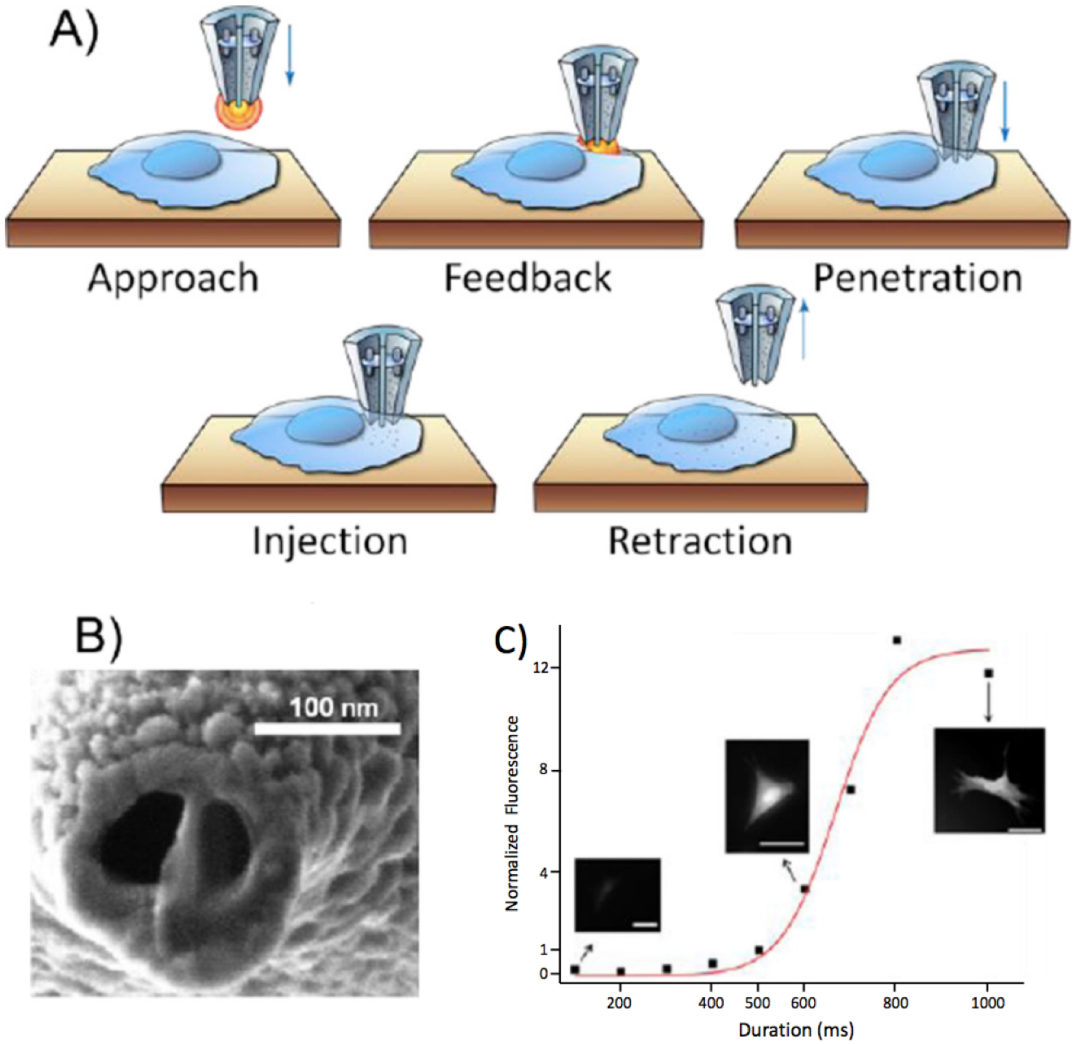

Figure A.1: (A) Schematic representation of cell-surface detection by a double-barrel nanopipette; (B) SEM image shows the gold-sputtered double-barrel nanopipette; (C) Injection of carboxyfluorescein into human fibroblasts. The fluorescence intensity was normalized to that measured at 500 ms. Applied voltage: 10 V, scale bars 50μm. The red curve is a sigmoidal fit to the experimental data points. (Reproduced from [9] with the permission of the Royal Society of Chemistry).

A detailed explanation of the feedback mechanism and the protocol can be found in Seger (2012) [9]. The post-injection long-term cell viability was monitored by injecting cells with carboxy-fluorescein succinimidyl ester into the cell and following the its morphology. Normal cell division was observed after 27 h of injection, with the daughter cells having normal cell morphology [9]. Future advances in the technology would enable a fully automated system for multiplex single-cell injections at the same time. Furthermore, interactions between different cell components, such as protein-RNA interactions, can be studied with this nanoinjection platform. RNA-binding proteins regulate the processes that RNAs are subjected to during and after transcription. Understanding protein-RNA interactions is crucial in illuminating their effect on the fate and function of RNA molecules. With our ongoing progress in monitoring the presence and concentration of proteins, nanopipettes can be deployed for real-time detection of complex protein interactions in the future. Further information on the transcriptome can be revealed by nanobiopsy and correlated with the behavior of the RNA-binding proteins [219].

Intracellular Tracking of Injected Components

Various molecules can be used to track a cell, such as genetically encoded fluorescent proteins [220], quantum dots [221] and fluorescent dyes. The ability to follow the process of the division of a single cell and how it transfers information to its daughter cells is key to advancing molecular biology and genomics. This ability would also be of benefit to developmental biology in understanding the conversion of a single cell into a full organism [222]. The ability to analyze the lineage history of cell populations reveal

information on developmental origin [223] and contribute to studies of disease transformation, seen in cancer, for instance [267]. Additionally, predicting the behavior and function of a single cell in complex tissue over time could contribute to pharmaceutical development and personalized medicine [268] by providing targets before pathogenesis. More efficient drug therapies, earlier intervention and recovery are sought to improve patient treatment and quality of life. Drug resistance during treatment is one of the major problems in many diseases, particularly in cancer [269, 270, 271, 179]. Resistance to a drug can even result in the resistance of cells to other pharmaceuticals [273], decreasing the chances of successful treatment. Therefore, it is crucial to examine the population of cells where resistance is developing in order to understand the molecular basis of drug resistance and to improve treatment outcome. In order to overcome drug resistance, researchers must understand how genomic changes are transferred from one cancer cell to another, including the perpetuation of drug resistance. Identified alterations can reveal genetic signatures of the development of drug resistance, leading to earlier intervention, modulation of therapy, and improved treatment outcomes.

A.2.2 Single-Cell Nanobiopsy Platform

Single-cell nanobiopsy platform was developed for continuous sampling of intracellular content from individual cells and has been described in detail elsewhere [201]. Because it is possible to extract a minute volume of material with a nano-size tip, it is possible to deploy this custom single-cell biopsy platform for extraction of cytoplasm from multiple locations in the same cell. The ability to map the subcellular distri-

bution of different biomolecules opens up new avenues of study; it is now possible to obtain information on cellular circuitry, neuronal growth, and network formation among cells, contributing to proteomics, genomics and diagnostics in several important fields of medicine and science.

A.2.2.1 Single Cell Immunoassay

Aside from carbohydrates and nucleic acids, proteins are one of the most common macromolecules in cells. Among these molecules, proteins are the most diverse molecules, playing a variety of biological roles: communication of information within and among cells, protection of cells against infection, catalysts for chemical reactions, and as structural components, to name but a few [274]. Therefore, there is great interest in quantifying, identifying and isolating proteins, in order to understand the plethora of unknown mechanisms in which they are involved. Conventionally, the methods of Lowry and Bradford were employed to quantify total protein content [275, 276]. However, these methods do not permit the identification of specific proteins involved in the processes of living cells. Subsequently, antibodies were utilized to identify specific proteins [277], and Southern blot, Northern blot and Western blot analyses were developed to detect DNA, RNA, and proteins, respectively. Western blot method adapted to detect single-cell proteins differentiated by molecular weight; this enabled the interrogation of more than 1000 cells in less than 4 h and multiplexed measurements of up to 11 proteins [278]. Single-cell Western blot, however, relies on separating the single-cell protein lysate using a polyacrylamide gel coating on a glass microscope slide,

which destroys the cell. Flow cytometry, microfluidics technologies, and surface methods such as ELISPOT were also studied as single-cell proteomic tools and have been extensively reviewed elsewhere [279]. The use of functionalized nanopipettes as a platform for label-free identification of biomolecules such as proteins has been strongly recommended. Also, protein-based recognition elements, such as antibodies and enzymes, can be functionalized in the sensing zone and further used for sensing of various molecules [280]. I will present a summary of sensing applications in the next sections of this Chapter. Functionalized nanopipettes are inserted into the single cell and used to monitor proteins in that cell. An antibody-labeled nanopipette shows excellent potential for the longitudinal interrogation of single cells. Implementation of this technology is on the cutting edge of advances in developing methods to combat human diseases. In addition to the proteomic approach, incorporating aspiration and sequencing of molecules from the nanopipette biopsy could identify significant disease-resistant variant genes. Therefore, the nanopipette can serve as a platform for integrated analyses of genomics, transcriptomics, proteomics and metabolomics of cells.

A.2.2.2 Genomics

The Human Genome Project pioneered research to identify the genetic entities behind conditions such as genetic disorders and drug resistance [281]. This research has become key in the process of drug discovery. However, drug discovery and diagnostics continue to present significant challenges. One difficulty lies in the heterogeneity of cells in complex tissues. The need to overcome this difficulty motivated the develop-

ment of useful tools for single-cell genomics and transcriptomics. These methods allow the examination of individual cells, circumventing the need to interpret pooled genetic information in population-based experiments, which will mask the earliest stages of change. Drug resistance, for example, can originate with mutations in an individual cell, which can then take over an entire population [198]. To overcome limits imposed by averaging out subpopulations in heterogeneous tissue, single-cell interrogation was proposed. Single-cell investigations that are destructive do not represent appropriate tools, because it is necessary to scrutinize genetic variation arising from the same cell over time. Some biological processes require monitoring of the same cell at multiple time points to understand the complete process. Fluorescence time-lapse microscopy was previously used to analyze gene circuit dynamics and heterogeneous cell behavior [282]. As an example, this technology was applied in embryonic stem cells to reveal the dynamics of the expression of pluripotency factor Nanog. Microscope-based detection of expression assessed different expression levels of the Nanog protein, demonstrating the interchangeable levels of Nanog-high and Nanog-low cells [283]. Our group deployed the nanopipette technology as a tool to aspirate the genomic information from single living cells and sequence the material with no destructive effect on the cell membrane [201]. The nanopipette platform enables non-destructive continuous longitudinal interrogation of single cells, and has the advantages to be non-destructive, label-free, single-cell system. Nanopipette also allows one to assess gene expression in subcellular compartments and organelles such as the cytoplasm, nucleus, and mitochondria. Figure 2 illustrates the genomics application of nanopipettes in aspiration and sequencing of mitochondrial

DNA.

A.2.2.3 Transcriptomic Analysis from Single Cell Aspiration

Extraction of molecules from a single cell by means of the nanobiopsy platform relies on electrowetting within a nanopipette. In brief, when an organic solution fills a nanopipette and the device is inserted into an aqueous solution, a liquid-liquid interface is created at the tip. Once voltage is applied between these two solutions, a force is produced at the interface which causes the solution to enter into or leave from the nanopipette [193, 284]. Under this condition, when a negative potential is applied, the solution moves toward the lumen of the nanopipette, and when a positive potential is applied, the solution moves to the outside of the nanopipette. In these interrogations, the amount of aspirated material from the cell compartment was estimated to be around 50 fL, or approximately 1% of the volume of a cell [201]. Our group integrated the nanopipette platform into a scanning ion conductance microscope (SICM) system that automatically positions the nanopipette above the cell of interest [9]. While in aqueous solution, the nanopipette is biased with a positive voltage to prevent the solution from flowing towards the lumen of the nanopipette. This voltage generates an ion current between the liquid-liquid interface, which can be used as the input of a feedback loop integrated with custom-built software. The software controls the movement of the nanopipette, continuing to approach the cell until a drop in the ionic current is detected, indicating the tip is at close proximity to the surface of the cell [9]. When a reduction of the electric current is detected, software stops movement in the direction of the

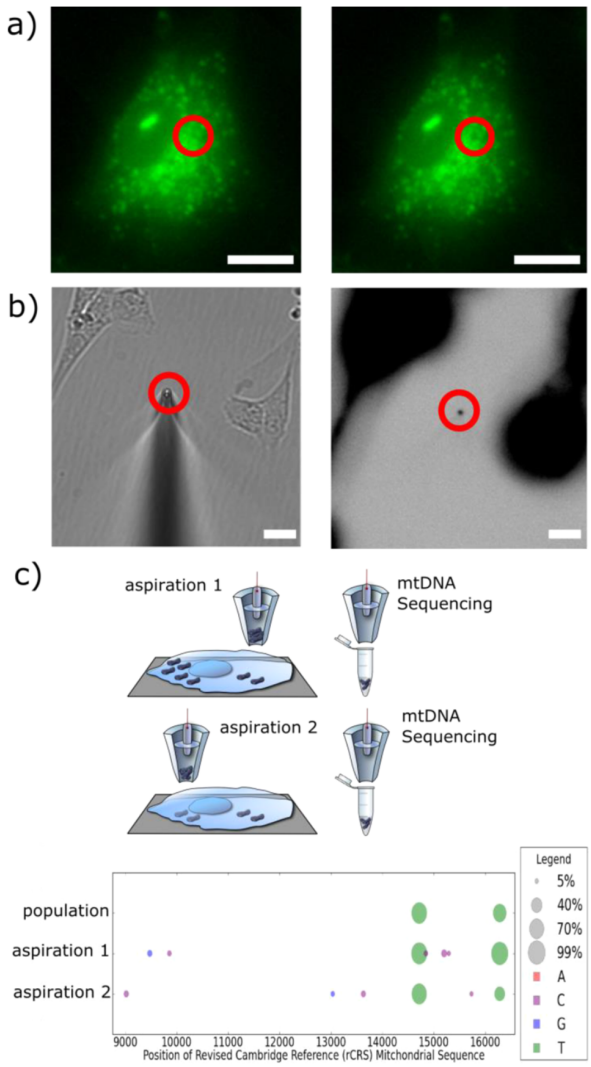

Figure A.2: Nanopipette isolation of mitochondrial DNA. (a) Fluorescent staining of human BJ fibroblast cells with MitoTracker Green before (right) and after nanobiopsy (left). Red circles show dark spot resulting of mitochondria removal. Scale bars 15 µm. (b) Nanopipette tip (red circle on left) used for mitochondria nanobiopsy in part a. Red circle on right indicates negative fluorescent of left panel showing fluorescence caused by nanopipette tip, indicating success of mitochondria nanobiopsy. Scale bars 15 µm. (c) Sequencing results demonstrate variable conservation of mitochondrial SNPs frequencies in aspirations. Heteroplasmic SNPs with estimated frequencies of 5% and 99% are displayed as circles whose area is proportional to observed frequency. Nucleotide of the variant is specified by color. A is red, C violet, G is blue and T is green. Variant 14713 A>T presents similar frequencies in aspirations and population. Variant 16278 C > T on the other hand, presents a greater variance of heteroplasmic frequencies in aspirations. Variants of low frequency were found in both aspirations but not in population.

cell and lowers the nanopipette at high speed (100 µm/s). This movement inserts the nanopipette into the cell membrane. The voltage applied to the nanopipette is then switched to 500 mV for 5 s, causing aspiration of cell cytoplasm into the nanopipette. Subsequently, a switch to 100 mV stops the influx, but does not induce the efflux of the aspirated content [11]. Nanopipettes fabricated from multiple-barreled capillaries allow the simultaneous injection of dye as molecules of biological interest are aspirated from the cell. Because of the small size of the device (approx. 50 nm), injury to cells from the nanopipette is minimal. Sequential delivery of multiple dyes has demonstrated the ability of the nanopipette platform to interrogate the single cell numerous times without fatally damaging the cell. Figure 3 shows the injection of multiple dyes into a single cell. Seger and collaborators demonstrated the ability of cells to survive for 27 h after the exposure [9]. These injections suggest the potential application of the nanopipette platform in multiple interrogations of the single cell, without lethal damage, which can be critical for development of single-cell drug resistance studies. Another study showed the use of nanopipettes to detect genes that were not previously described in the body of neurons by finding the compartmentalization of mRNA molecules in different parts of neurons [285]. For the mRNA molecule to be interrogated, it must first be sequenced.

A.2.2.4 Nanogenomics

The nanopipette can also be employed to aspirate cell contents from the same single cell multiple times during its lifetime to study molecular dynamics. This platform was previously validated to isolate molecules such as RNA for cDNA synthesis

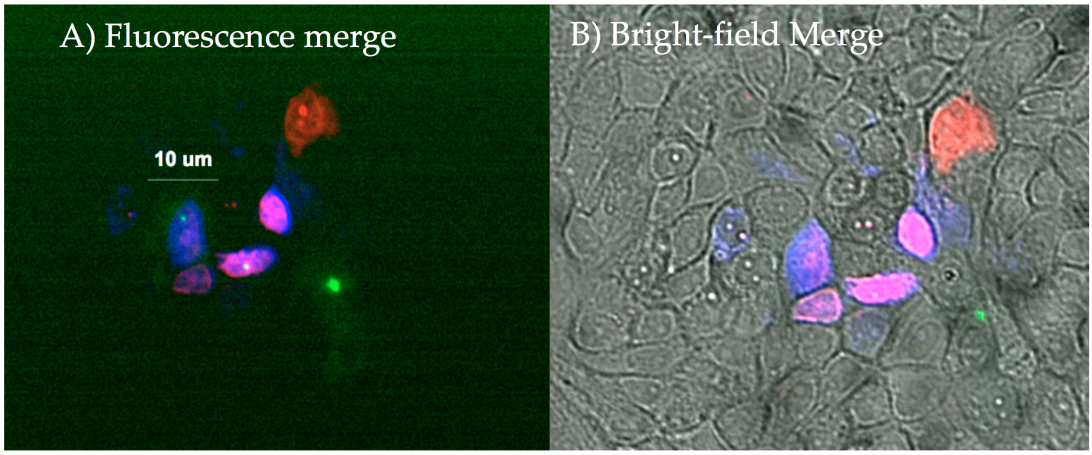

Figure A.3: (A) Fluorescence; (B) Bright-field merges show injections of green fluorescent protein rhodamine, and mitotracker orange into the cells. GFP: green channel; mitotracker orange: blue channel, rhodamine: red channel. Cells stained purple are a mix of blue (mitotracker) and red (rhodamine) channels. One cell at the center can be seen with GFP, mitotracker and rhodamine fluorescence, indicating three nanopipette interrogations. GFP was the first component to be injected into the cell, however it did not diffuse well into the cell, probably due to protein viscosity. After GFP, mitochondria-staining dye mitotracker orange was introduced. Rhodamine was injected as the third component into the group of cells.

and qPCR. Our group became one only a few groups to perform Next Generation Sequencing (NGS) from the species extracted with nanopipettes. Nashimoto's research group has shown device automation in the ZYX axis for isolation of mRNA molecules [286]. Guillaume-Gentil has demonstrated the identification of metabolites and enzymes using atomic force microscopy and also validated mRNA aspiration using qPCR [287]. However, analytical techniques such as NMR and MS spectrometry for the detection of single-cell molecules are still limited. In 2007, Luo and Li reported on the identification of 12C/13C-dansyl labeled metabolites by means of MALDI-MS in a minimum of 100 cells [288]. The group was able to detect subpopulations of heterogeneous tissue, but technical limitations of the method did not allow single-cell resolution. Guillaume-Gentil also reported the utilization of atomic force for aspiration and detection of mRNA molecules [289]. Cao and collaborators demonstrated longitudinal interrogation of single cells, sampling GFP and RFP transcripts from cells [266]. These techniques, on the one hand, relied on the observation of aspiration by fluorescence or qPCR amplification. Genes of interest, on the other hand, are not always tagged with fluorescent protein to identify protein localization. Also, not all RNA molecules involved in genetic mechanisms are expressed as proteins. However, it is not rare that all the genes of a cell must be interrogated. To successfully identify the highest possible number of genes involved in drug resistance, interrogation of cells can only be accomplished using next-generation DNA sequencing platforms. To show that nanopipettes did not affect in the function of cells upon piercing the cell membrane, human BJ fibroblasts were treated with Ca2+ agent Fluo4 AM, and fluorescent microscopy was used to show the

localization of Ca2+ ions before, during and after nanopipette biopsy [201]. Optic microscopy images showed that the procedure was minimally invasive, generating only a small change of Ca2+ during nanobiopsy. The cell recovered a few seconds after the process, reaching Ca2+ concentrations that matched pre-aspiration levels. By contrast, Actis et al. demonstrated that micropipette aspiration caused dramatic changes in the concentration of Ca2+ ions in the cell [201]. The low interference of nanopipettes results from the minimal interaction of the nanopipette with the surface membrane of the cell, contrasted with the highest surface of communication and damage demonstrated by micropipettes, indicating a better outcome for nanopipette interrogation compared to micropipette interrogation. It is important to note that nanopipette aspiration is based on a voltage-controlled influx of material and not adsorption of molecules to the walls of nanopipette. PCR amplification of DNA templates was not observed if negative voltage was not applied to the nanopipette during single-cell interrogation and when aspiration was performed in the bulk solution. This is the critical element that differentiates the nanobiopsy technology from AFM-based platforms. Both Wickramasinghe's and Osada's groups used AFM probes to extract RNA from cells in culture, either based on physisorption or hybridization of complementary RNA immobilized onto the probe [224, 225]. We foresee that the use of nanopipettes to aspirate limited copies of mitochondrial DNA from a living cell might provide the basis for less invasive and more accurate monitoring of disease progression. The potential of nanobiopsy is also such that the foundation can be established for the development of new classes of drugs to attenuate diseases as diverse as Parkinson's and Alzheimer's Disease. The nanopipette

can be used as a platform for clinical management of cancer research, elucidating the role of heterogeneity in primary tumor tissues and systemically identifying critical parameters in disease progression and potential metastatic states [226, 227]. By combining the nanopipette platform with downstream sequencing implementation, gene expression inside single cells can be longitudinally investigated, and the effect of drug mechanisms on mutation-selection can be better examined. The nanopipette platform also allows subcellular interrogation. By using different dyes in the cellular nucleus or by staining the cytoplasm, enabling the visual isolation of the nucleus, it is possible to target the two compartments differentially. The following pictures in Figure 4 show cells stained with mitochondria dye (mitotracker orange).

The chromosomal region can be distinguished from the cytoplasmic by observing the white granulocytes that correspond to the interaction of mitochondrial proteins with the dye. The nucleus is depicted as circular black orifices without mitochondria. The nanopipette was inserted into the dark orifice, corresponding to the cellular nucleus. To control for the downstream sequencing process of nanopipette aspiration, we implemented the addition of External RNA Controls Consortium (ERCC) spike-in controls with samples collected from the cells. ERCC controls are a set of RNA standards for use in microarray, qPCR and sequencing applications [228]. These molecules are artificial poly-adenylated RNA, used in library preparation protocols before cDNA synthesis. The ERCC internal control is designed so that increased variability is detected as the number of reads decrease. However, the nanopipette biopsy was still able to identify reads mapping to the human genome [285]. Therefore, coupling the nanopipette plat-

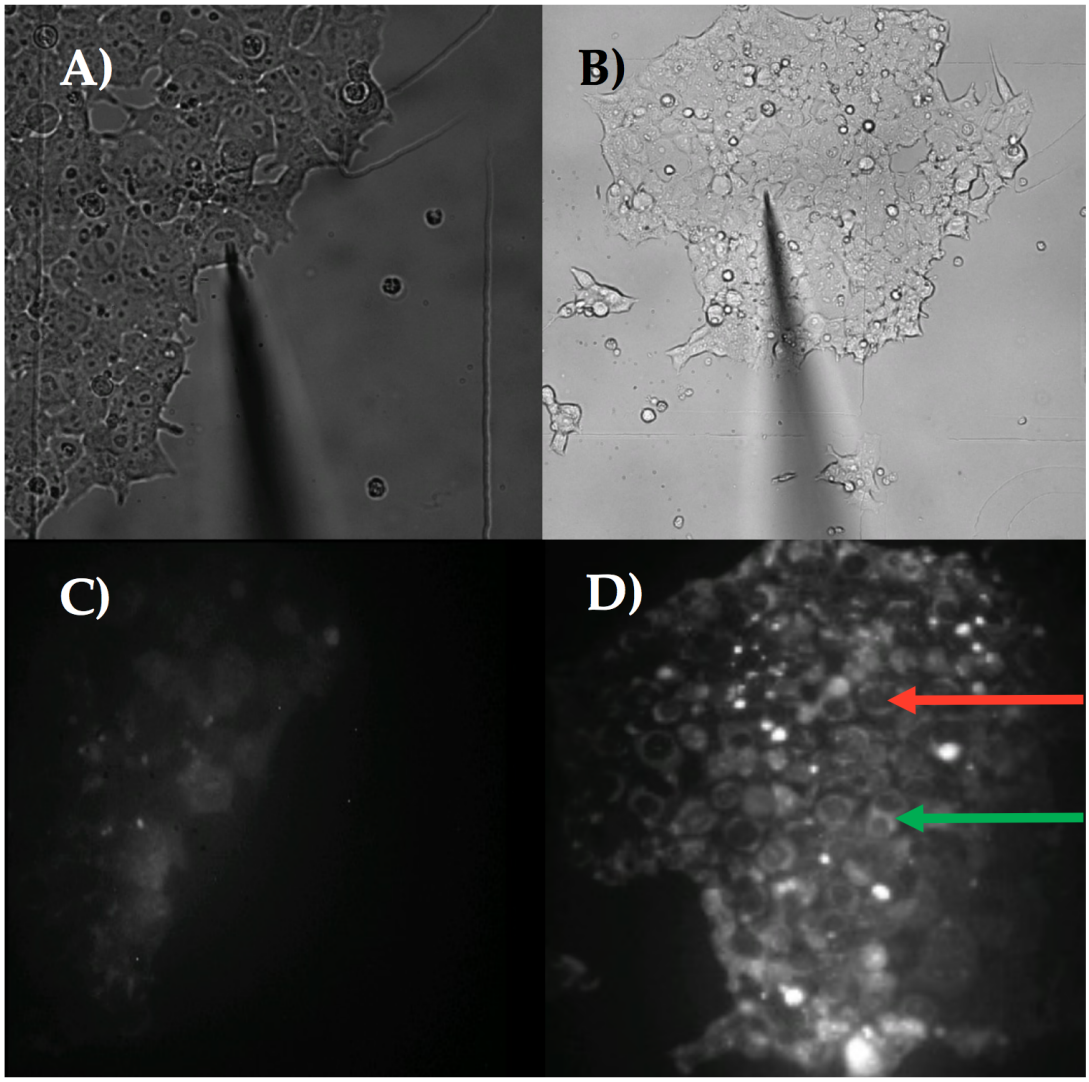

Figure A.4: Aspiration of nuclear content by Nanopipette. (A) Nanopipette is placed on top of MCF-7 cell; (B) Nanopipette is placed on top of a different MCF-7 cell; (C) Fluorescence corresponding to mitotracker orange staining of cells depicted in (A); (D) Fluorescence corresponding to mitotracker orange staining of cells depicted in (B). Nuclear region is visualized by pattern of staining with the mitotracker dye. In (D) red arrow points to dark compartment, corresponding to one nucleus. Green arros shows one cytoplasmic region. Nanopipette was inserted into the nucleus, as seen in (B). Nuclear content was aspirated and transferred to the cDNA synthesis master mix, followed by sequencing using the Illumina MiSeq instrument.

form with the sequencing of mRNA molecules showed the ability of the nanopipette platform to successfully identify low-abundant molecules in the context of gene expression, a capability essential for single-cell interrogation. ERCC spike-ins were used to show the ability of nanopipettes to isolate cellular RNA molecules for sequencing. Reads used were those that mapped to at least one spot in the human genome, as described by Actis (2013) [201] and Toth (2018) [285].

After separation of the ERCC counts from reads proceeding from cellular content, reads mapped to the human reference genome were plotted as Principal Component Analysis (PCA) results of cellular expression, showing the clustering pattern of the nuclear aspirations of single cells. The PCA of gene expression in the nuclear nanobiopsy samples, using both non-processed and pre-processed gene counts, are shown in Figure 6. The sequencing reads were aligned against the human reference genome using the STAR aligner, and the HTSeq package was used to count the number of mapped reads. Using the limit of detection (10 reads per detected transcript), reads were input to DESeq2 (HL = HeLa transcriptome library; MBL = MDA-MB-231 transcriptome library; NL = iCell neuron library; MCL = MCF-7 transcriptome library). Figure 6A shows the PCA of gene expression in the nanobiopsy samples. Libraries MBL1, MBL9, MBL12, and MBL14 were considered outliers and removed from downstream analysis. The Figure 6B graphs are plotted from PCA runs with reads log-transformation, with the aim of mitigating the variation effect of highly expressed genes or any biases possibly introduced during the cell nanobiopsy procedure, library preparation or sequencing run. In Figure 6C we focused on the areas from Figure 6A that have more clustering

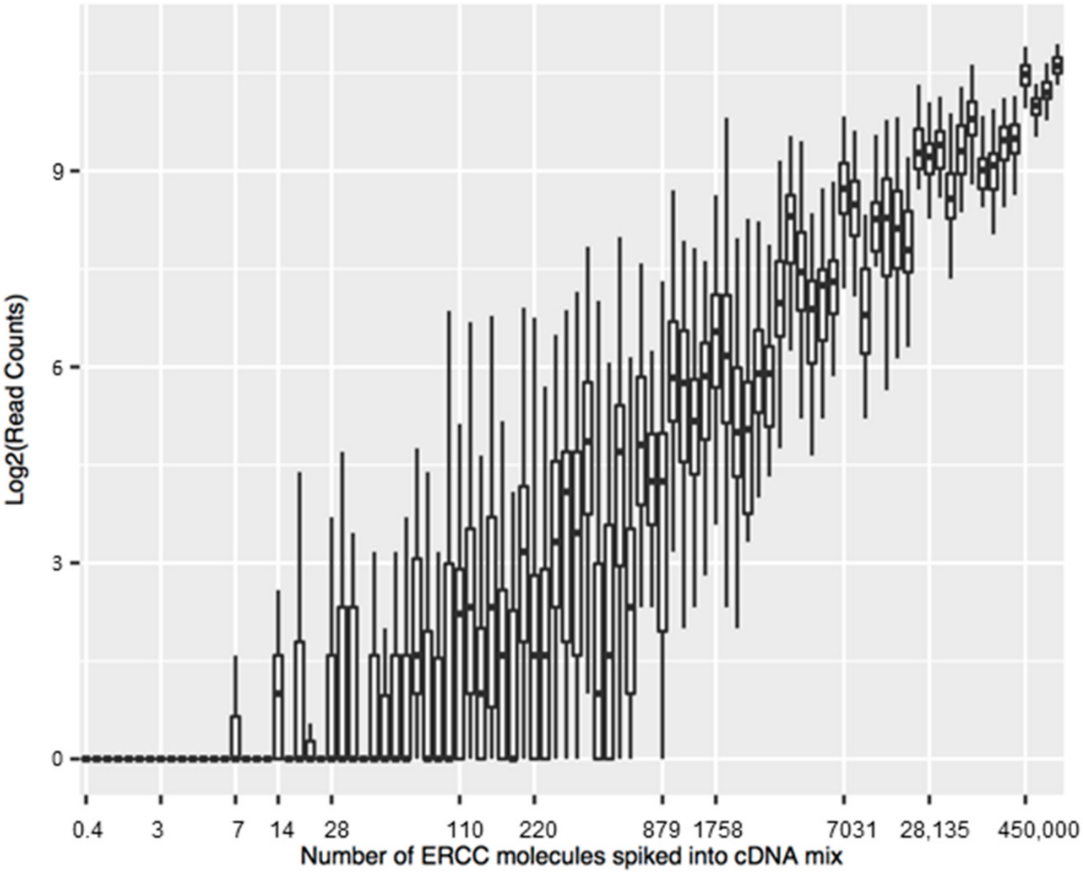

Figure A.5: Limit of detection of ERCC RNA molecules as a control for nanopipette biopsies. Content from nuclear nanobiopsy was transferred to cDNA mix (containing 0.5 µL of ERCC mixture at a 1:10,000 dilution) for synthesis of cDNA (containing 0.5 µL of ERCC mixture at a 1:10,000 dilution) to reverse transcribe the RNAs followed by cDNA sequencing. Sequencing reads were mapped to the ERCC reference pseudo-genome. Numbe of RNAs followed by DNA sequencing. The sequencing reads were mapped to the ERCC reference pseudo-genome. The number of transcripts were counted using the HTSeq package and plotted as a function of the number of ERCC transcripts (ERCC concentration × volume × dilution factor). The estimated intersect of the ERCC curve with the X axis was between 7 and 220, which represents at least one detected ERCC transcript. The threshold for detected transcripts was chosen to be 10 for subsequent analysis.

structure for better visualization. Furthermore, Figure 6D was plotted to give a closer look to the clustered areas from Figure 6C. Figure 6C,D illustrate the expression profile of the four cell types are different from one another.

It was not clear to what extent the MDA-MB-231 cells and MCF-7 cells were distinguishable using PCA in Figure 6. Therefore, we plotted the MDA-MB-231 cells and MCF-7 separately in (Figure 7A–C). Figure 7B represents focused areas of Figure 7A, Figure 7C represents focused areas of Figure 7B, for more clustering structure. Figure 7E represents focused areas of Figure 7D for more clustering structure. These results suggested that, although the number of detected reads is small per sequenced library, nanopipette technology detects the similarities of same-cell types based on the gene transcription pattern. Figure 7D–F support the conclusion of MDA-MB-231 vs. MCF-7 comparison. Figure 7E represents focused areas of Figure 7D, Figure 7F represents focused areas of Figure 7E, with more clustering structure.

To determine the identity of more abundant genes in the MDA-MB-231 and MCF-7 cells, we extracted RefSeq IDs with more than 200 reads in at least one of the 39 sequenced libraries from the dataset, and checked the presence of the genes in both MDA-MB-231 and MCF-7 cells. Table 1 represents the ability of nanobiopsy to resolve the identity of a cell type by detecting highly abundant transcripts associated with ubiquitous biological processes.

As an example, genes associated with glucose metabolism (UGP2, ENO1), ribosomal protein synthesis (RPLP0, EEF1A1, NPM1), protein folding (HSP90AA1), protein degradation (POMP), DNA binding (H3F3B), and drug resistance by cancer

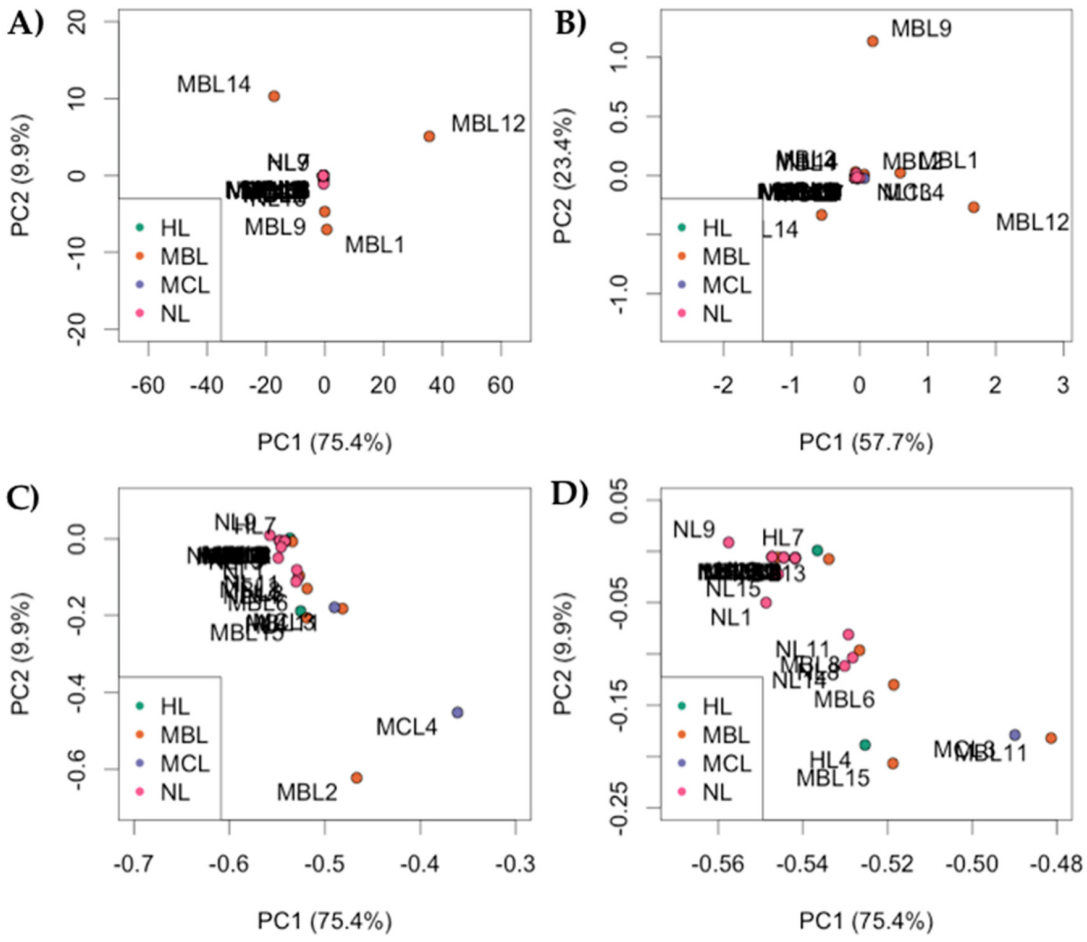

Figure A.6: Principal Component Analysis of gene expression in the nuclear nanobiopsy samples. (A) Raw data input to DESeq2; (B) DESeq2 run with log-normalized reads; (C) Resolution of clustering after removal of the MBL1, MBL9, MBL12 and MBL12 libraries as outliers; (D) Resolution of clustering excluding sequencing libraries MBL2 and MBL4.

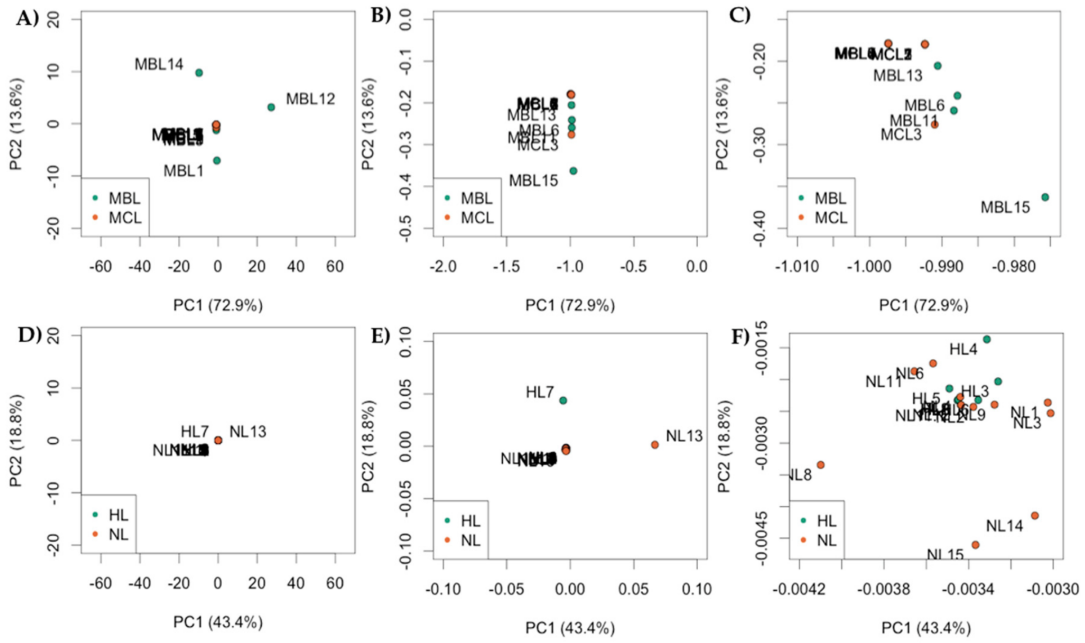

Figure A.7: Principal Component Analysis of gene expression comparing two cell types at a time. (A–C) comparison of MDA-MB-231 and MCF-7 libraries cluster separately by cell type, seen as a trend in comparison of MDA-MB-231 and MCF-7 libraries cluster separately by cell type, seen as a trend in which same-cell type libraries cluster closer to each other; (D–F) comparison of HeLa vs. iCell which same-cell type libraries cluster closer to each other; (D–F) comparison of HeLa vs. iCell Neurons cells. Libraries cluster separately by cell type. (Pourmand Lab, Personal Communication, 2018).

Table A.1: Genes detected commonly in MDA-MB-231 and7 MCF-7 cells. Libraries that had at least one gene with 200 reads were qualified for mapping using RefSeq IDs. The genes displayed in the table were detected both in the MDA-MB-231 and MCF-7 cells.

RefSeq Accession	**Symbol**	**Gene Name**
NM_001001521	UGP2	UDP-Glucose Pyrophosphorylase
NM_001002	RPLP0	Ribosomal Protein Lateral Stalk Subunit P0
NM_001017963	HSP90AA1	Heat Shock Protein 90 α family class A member 1
NM_001201483	ENO1	Enolase 1
NM_001402	EEF1A1	Eukaryotic Translation Elongation Factor 1 α 1
NM_001699	AXL	AXL Receptor Tyrosine Kinase
NM_002520	NPM1	nucleophosmin 1
NM_005324_mRNA	H3F3B	H3 Histone, Family 3B
NM_015932_mRNA	POMP	Proteasome Maturation Protein

cells (AXL) [229, 230, 231, 232, 233, 234]. More specifically, genes ENO1, H3F3B and HSP90AA1 are important cancer drivers in human cells.

A.3 Monitoring Intracellular Components by Nanopipette Sensing

The identification and quantification of molecules in the single cell play a crucial role in diagnostics and fundamental molecular biology. The ability to dynamically monitor the presence and amount of any molecule and/or biomarkers in the single cell aids in understanding the relationship of these molecules to several diseases, and contributes to drug discovery research. The sensing region or the nanopipette tip surface responds to changes in the ionic current flowing through the pore, which can be brought about by electrostatic, biotin-streptavidin, or antibody-antigen interactions. Specific antigen-antibody interaction changed the current amplitude and showed reasonable promise for future applications in biomolecular diagnosis. The successful implementation of nanopipette technology in biosensing enabled the identification of a variety of molecules from glucose to proteins. In the following section, we review the commonly used immobilization techniques, techniques to generate signal, and recognition probes on the nanopipette. Specific examples from the literature are given subsequently.

A.3.1 Layer-by-Layer (LbL) Immobilization of Recognition Elements

The idea of running voltage through the nanopore and using the resultant current as feedback originated in 2002 [235]. Beginning in 2004, numerous groups began using nanopipettes as a transport system for metals and small molecules [212, 236, 237]. In 2006, the idea of attaching polymers to the nanopipette in order to increase feedback responses came to fruition, when it was shown that the surface charge could be manipulated by coating the pore with Poly-L-Lysine (PLL). The attachment of PLL was confirmed by the current rectification observed when running voltage through the tip of a coated nanopipette into a solution containing 25 mM KCl [238]. This feedback mechanism, which gave scientists the ability to confirm that a polymer was attached, allowed for further functionalization of nanopipettes in the form of layer-by-layer (LbL) assemblies. In 2009, Umehara et al. posited that the attachment of specific probes to the nanopipette pore could lead to label-free, quantitative sensing of small molecules, proteins, and/or antigens [280]. That paper was the first to detail an approach for LbL assembly that allowed the detection of specific proteins in solution using antibodies. First, PLL was coated and baked onto the bare surface of the nanopipette [239]. 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and N-hydroxysulfosuccinimide (NHS) were then deposited to create a link between PLL and the subsequent layer [240]. Protein A/G was then conjugated to the NHS/EDC linker [241]. Finally, IgG was immobilized to protein A/G on the tip of the nanopipette [242]. During the same year, this type of LbL assembly was patented [243].

A.3.2 Electrochemical Techniques Used for Analysis

Nanopipettes have not always been functionalized into biosensors as they are today. Beginning in 2002, the Korchev and Klenerman labs began using nanopipettes as a new SICM probe for cellular structures and substructures [**?**, 237, 244, 245]. In this setup, there is a reference electrode (RE) inside the nanopipette and a working electrode (WE) in the solution. Voltage is applied between the two electrodes, and the resultant current is used to gauge how far away the nanopipette is from the structure [233]. The nanopipette is attached to a piezo stage, which controls its xyz movement. Current is kept constant as the nanopipette is moved along the xy plane by adjusting its z position. By using this current feedback system, researchers are able to provide high-resolution imaging. Once a topological map has been drawn, the computer-controlled piezo stage can be used to precisely position the nanopipette over a feature of interest, such as an ion channel. Then, the current feedback system is switched off, allowing the nanopipette to be lowered to the surface of the channel. At that point, suction is applied, forming a giga-seal, which allows noise-reduced patch-clamp recordings to be made [246]. In addition to high-resolution imaging and patch-clamp recording, nanopipettes have been explored as vectors for molecule delivery. At approximately this time, research began using nanopipettes for SICM; the labs cited above as well as others began to use nanopipettes to capture and transport DNA, metal ions, and other molecules [236, 237, 247, 248, 249]. In one example, carbon nanotubes were filled with iron; a current was run through the tube, and iron flowed out of the tube and was deposited on a surface

[236]. Elsewhere, the same schematic was used as described above, where a reference electrode (RE) is placed inside the nanopipette while the working electrode (WE) is in a buffer solution [237, 247, 248, 249]. In these cases, a potential waveform is applied between the electrodes that influences the electroosmotic flow (EOF), electrophoretic flow, and dielectrophoresis of DNA, metal ions, or proteins in solution. Klenerman's group has reported variations observed when testing conditions are kept constant [247], suggesting that the electrical field and gradient inside the nanopipette is extremely sensitive to the geometry of the nanopipette. In 2004, the Stanford Genome Technology Center began to use a nanopipette as an electrochemical biosensor [200]. Unlike previous works, the WE was placed inside the nanopipette, and the RE was placed in the bath solution. When voltage was applied between the two electrodes and the resultant current was measured, the Stanford group was able to observe translocation of DNA labelled with gold nanoparticles (DNA-AuNPs) by observing short bursts of current reduction when the DNA-AuNPs translocated into or out of the nanopipette. The same group began exploring the behaviors of functionalized nanopipettes. These experiments, which largely used the same electrochemical techniques, focused on current rectification of the nanopipette system at various applied voltages [238]. It was observed that coating nanopipettes with PLL could modulate current at particular applied voltages and could amplify signals produced under certain conditions. This led to the development of a layer-by-layer assembly on the inner pore of the nanopipette [280, 243], with antibodies attached to its outermost layer, making it specific to a particular antigen. By applying a constant voltage, the researchers were able to see changes in current when the antigen

was added to the solution. Variations of this assay exist in several forms. The Long lab, for example, has observed the ability to differentiate alpha-fetal protein (AFP) from AFP bound to its conjugate antibody (AFP-anti-AFP) [250]. In their experiments, they measured translocation events by observing the change in current as the protein or the antibody-bound protein transverses the nanopipette's pore. It was shown that when the antibody binds to AFP the translocation is longer and more complete, causing a larger reduction in current. Today, the basic schematic for a nanopipette is to have two electrodes, one inside the nanopipette and one in the surrounding solution. A waveform is then applied through the electrodes and the resultant current is observed. The current can be modulated and changed based upon how the pipette has been functionalized or on the target for observation in the solution.

A.3.3 Recognition Element Selection for Immobilization on Nanopipettes

The recognition element is one of the major factors affecting nanosensor performance. The specificity of a nanopipette-based biosensor is restricted by the molecule deployed as the recognition element. Receptors, enzymes, antibodies, and nucleic acids can be employed in the sensor design to recognize the target of interest [251]. Initially, affinity reagents, such as monoclonal antibodies and enzymes obtained from living systems, were deployed in sensor construction. However, numerous concerns with monoclonal antibodies, such as reproducibility of the clone, high production costs, stability, and cross-reactivity, led to the development of oligonucleotide-based molecules for recognition [252]. Nucleic acid aptamers are attractive alternatives to protein-based

recognition probes with lower cost, longer shelf life, and less batch-to-batch variation derived from the chemical production process [253]. Like antibody detection, aptameric detection takes advantage of specific binding between its conjugate and itself, which causes a change in current rectification that occurs as the pore of the nanopipette is blocked, and/or a change in the surface charge. The selection criteria for the recognition molecule of nanopipette-based sensing largely depends on the application area and the analyte. Using antibodies on the outermost layer can provide specific sensors; other molecules, such as aptamers [128] or enzymes [16], can readily be used as other recognition molecules for detection. Enzymes can also be used in a secondary detection method. For example, glucose oxidase was attached to the outermost layer of a nanopipette pore, causing the oxidation of -D-glucose to D-gluconic acid. This led to a drop in pH, which caused a measurable change in current rectification [16].

A.3.4 Specific Examples from the Literature

A.3.4.1 Glucose

The differences between glucose levels of individual cells may be indicative of diseases such as cancer [255]. These changes can further assist in the identification of abnormal cells. After identification, these cells can be labeled and followed over the course of treatment. For example, increases in glucose levels were observed in the metastatic breast cancer lines MDA-MB-231 and MCF7 compared to nonmalignant cells [16]. This increase in glucose consumption contributed to the tumor cells' rapid growth and proliferation, which requires increased metabolic activity. It was also speculated that altered

glucose metabolism can result in metastasis and resistance to chemotherapeutic drugs [256]. Therefore, it is essential to monitor the metabolic activity of the single cell, not only for identification, but for the ability to study the transformation of single cancer cells in the heterogeneous cell population. We have used the nanopipette as a platform to immobilize glucose oxidase (GOx) for real-time intracellular glucose detection and have monitored the changes in impedance [16] (Figure 8).

A direct relation between impedance change and glucose concentration was observed and used to create a calibration curve. Notably, the surface chemistry developed for GO can further be employed for the attachment of various enzymes and applied to detect different substrates in the cells.

A.4 Conclusions and Future Perspectives

In the last decade, much effort has been concentrated on applications of nanopipettes as single-cell surgical tools both for injection and for aspiration of various materials. The combination of nanopipettes as surgical tools and selective sensing tools enabled detection of biologically relevant compounds in single living cells and in well-defined regions of the cell compartment. The advances in nanopipette technology have also benefited various areas of molecular biology research, including but not limited to proteomics and genomics. In the future, the technology presented here can be applied to automated cellular collection systems, which could allow the researcher to perform several tasks at the same time. There is significant interest in producing nanopipettes with precisely

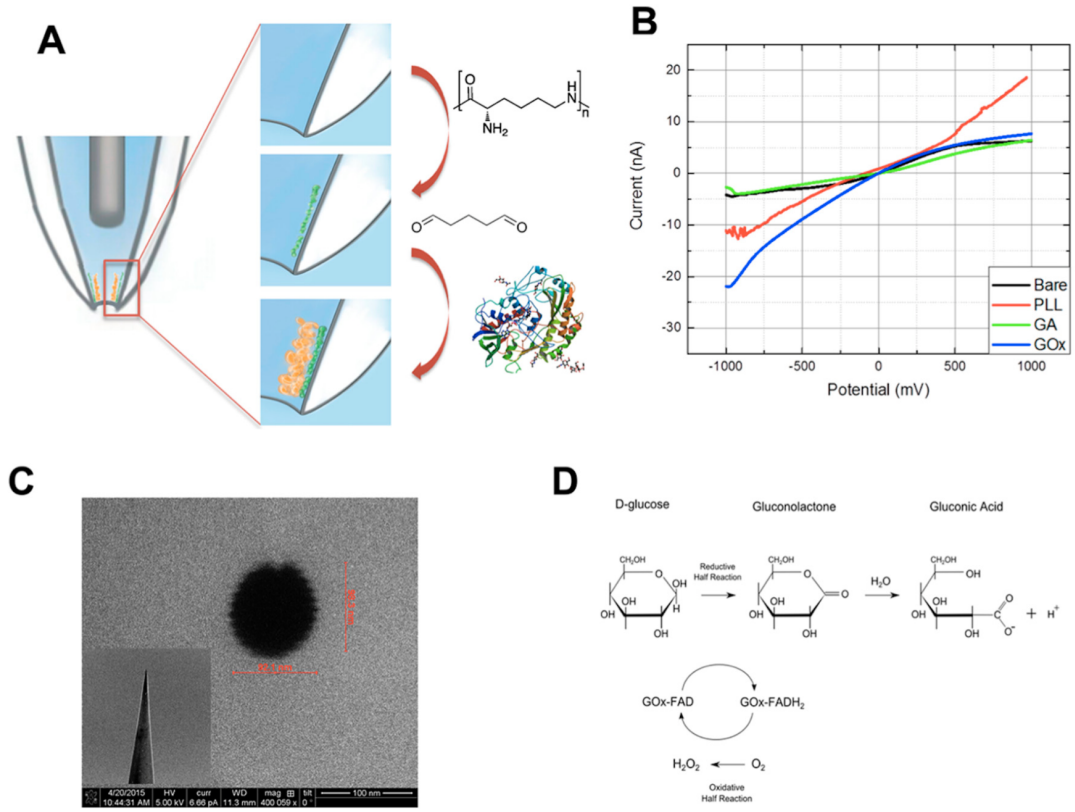

Figure A.8: Representative schematic showing the steps of glucose oxidase immobilization to the surface of the nanopipette tip. First, PLL is coated on the surface. Then, gluteraldehyde treatment occurs to cross-link the glucose oxidase to the PLL-coated surface; (B) After each step of immobilization, the changes were characterized electrochemically. 10 mM PBS (pH 7) was used as the supporting electrode; (C) Nanopipette tip imaged by SEM. Tip geometry is displayed in the inset; (D) Enzymatic process for conversion of glucose into hydrogen peroxide and gluconic acid. (Reprinted with the permission from [16]. Copyright (2018) American Chemical Society).

identical geometrical parameters, such as size and shape of the tip. Post-processing of nanoparticle production can help fine-tune these parameters by applying different approaches. Development of precise tip parameters is of scientific interest not only for sensing applications but would also be of interest to researchers dealing with dielectric etching. We believe that with increased attention to nanopores, nanogating, and ion channels, nanopipettes will find even broader application in a variety of fields, from electrophysiological to medical research, and will become a fundamental tool for single-cell studies.