# The Influence of Single Nucleotide Polymorphisms and Epigenetics on HIV Disease Progression

by

**Jenna Malinauskas**

BS in Microbiology, University of Wisconsin La Crosse, 2016

Submitted to the Graduate Faculty of

the Department of Infectious Diseases and Microbiology

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

**Jenna Malinauskas**

It was defended on

December 10, 2018

and approved by

**Thesis Advisor:**
Jeremy Martinson, DPhil
Assistant Professor
Infectious Diseases and Microbiology
Graduate School of Public Health
University of Pittsburgh

**Committee Member:**
Robert Mailliard, Ph.D
Assistant Professor
Infectious Diseases and Microbiology
Graduate School of Public Health
University of Pittsburgh

**Committee Member:**
Bernard Macatangay, MD
Assistant Professor of Medicine
Division of Infectious Diseases, Department of Medicine
School of Medicine
University of Pittsburgh

Jeremy Martinson, DPhil

**The Influence of Single Nucleotide Polymorphisms and Epigenetics on HIV Disease Progression**

Jenna Malinauskas, MS

University of Pittsburgh, 2018

## Abstract

Human immunodeficiency virus (HIV) is one of the most significant public health issues globally, however; little is known about the factors contributing to the rate of AIDS progression. Due to antiretroviral therapy, disease progression classification can no longer be attained, therefore; retrospective samples from the Multicenter AIDS Cohort Study (MACS) are used to study host genetic factors contributing to progression though studying Nonprogressors (NPs) and Progressors (PRs). The genetic response to HIV infection can be influenced by polymorphisms, or epigenetic factors such as chromatin accessibility. Using previously acquired microarray data, we observe multiple genes differentially regulated when comparing immature dendritic cells (iDCs) from NPs and PRs, and recent studies by Rappocciolo et al, 2014 demonstrate that iDCs from NPs have inefficient *trans*-infection due to reduced cellular cholesterol. I hypothesize that sequence variation and/or epigenetic regulation may be influencing these processes and contributing to slower disease progression.

Next generation sequencing was used to determine whether SNPs in the promoter regions of *palladin*, contribute to differential gene expression. Assay for transposase chromatin accessibility (ATAC-seq) was validated for use on MACS samples to determine if chromatin accessibility differs between NPs and PRs.

Promoter region sequencing results demonstrate a variant within *palladin* with an allele frequency of 0.929 in the NP cohort near the promoter region of isoform 4. The variant is absent in the PR cohort and is located within the transcription factor binding site for RB1. RB1 controls heterochromatin structure contributing to gene expression, therefore; a variant within the binding site may influence isoform 4 expression in NPs.

ATAC-sequencing results indicate successful sequencing of healthy donor iDCs using the Omni-ATAC-seq protocol described by Corces, et al, 2018. Library complexity, sequencing depth and coverage were acceptable when analyzed with ATACseqQC. These results suggest that the Omni-ATAC-seq method can successfully be used on MACS cohort samples to determine differences in chromatin accessibility between NPs and PRs.

**Public Health Significance**: Understanding how host genetics influence disease progression to AIDS is important for the development of new therapies. Additionally, we are discovering novel innate mechanisms of infectious disease control that may be applicable to viral infections broadly.

## Table of Contents

**Preface** ..... x

**1.0 Introduction** ..... 1

**1.1 Natural History of HIV Infection** ..... 1

**1.2 HIV Pathogenesis** ..... 2

**1.3 Host Genetics and Progression to AIDS** ..... 3

**1.4 Multicenter AIDS Cohort Study Participants** ..... 4

**1.5 *Cis* vs. *Trans* Infection** ..... 5

**1.6 Epigenetic Regulation of Gene Expression** ..... 6

**1.7 Genes of Interest** ..... 8

**1.7.1 Palladin (PALLD)** ..... 9

**2.0 Statement of Project** ..... 11

**2.1 Aim #1** ..... 11

**2.2 Aim #2** ..... 11

**3.0 Research Design** ..... 12

**3.1 Amplicon Sequencing** ..... 12

**3.2 TaqMan® Genotyping** ..... 17

**3.3 ATAC-Sequencing Cell Culture** ..... 18

**3.4 Transposition Reaction** ..... 18

**3.5 ATAC Library Preparation and Sequencing** ..... 19

**3.6 ATAC-Sequencing Data Analysis** ..... 21

**4.0 Results** ..... 22

**4.1 rs6820223 in *palladin* is Present in All NPs and Absent in PRs ..... 22**

**4.2 *palladin* Variant is in the RB1 Transcription Factor Binding Site Near Isoform 4 Promoter Region ..... 23**

**4.3 rs6820223 Allele Frequency Was Not Replicated in a Larger MACS Cohort ..... 23**

**4.4 ATAC-Sequencing Library Complexity and Mapping ..... 24**

**5.0 Discussion ..... 27**

**5.1 Transcriptome Analysis Determines Palladin Expression in Nonprogressors with Inefficient *Trans* Infection is Significantly Higher than in Progressors ..... 27**

**5.1.1 Palladin is a Cytoskeletal Scaffold Protein with Drastic Effects on Cellular Function ..... 28**

**5.2 Amplicon Sequencing Detects Variant in the Third TSS of palladin with a High Allele Frequency in NPs ..... 29**

**5.3 Allele Frequency of rs6820223 in Modest Cohort was not Replicated in a Larger Cohort Analysis ..... 29**

**5.4 The Omni-ATAC-Sequencing Method Successfully Removes Mitochondrial Contamination and can be Used to Determine Chromatin Accessibility in NPs and PRs ..... 30**

**5.5 Experimental Limitations ..... 32**

**5.6 Conclusion and Future Directions ..... 32**

**Bibliography ..... 34**

## List of Tables

Table 1. Primer Sequences for *palladin* Transcriptional Start Sites ..... 12

Table 2. PCR Conditions for Transcriptional Start Site 1 of *palladin* ..... 13

Table 3. PCR Conditions for Transcriptional Start Site 2 of *palladin* ..... 13

Table 4. PCR Conditions for Transcriptional Start Site 3 of *palladin* ..... 13

Table 5. *palladin* PCR Reaction Components ..... 14

Table 6. Library Concentration and Molarity for Amplicon Sequencing ..... 16

Table 7. TaqMan® Genotyping Assay qPCR Components ..... 17

Table 8. TaqMan® Genotyping Assay qPCR Cycling Conditions ..... 17

Table 9. Library Concentrations via Tapestation 2200 ..... 19

Table 10. Pre-Amplification PCR Master Mix ..... 20

Table 11. Nextera Index Kit PCR Primers ..... 20

Table 12. Nextera Index Kit Index 1 (i7) Adapter ..... 20

Table 13. Nextera Index Kit Index 2 (i5) Adapters ..... 20

Table 14. Reagents for qPCR to Determine Additional Cycles ..... 20

Table 15. Allele Frequency Table of *palladin* rs6820223 in NPs and PRs ..... 24

Table 16. Summary Statistics of the 3 ATAC-seq Libraries ..... 25

## List of Figures

Figure 1. Differentially expressed genes between MACS NPs and PRs. ..... 9

Figure 2. *palladin* Transcriptional Start Site 1 Amplicons of 16 NPs, PRs and SN in Triplicate 14

Figure 3. Representative IGV Image of NP 4 Near Transcriptional Start Site 3 ..... 22

Figure 4. Allelic Discrimination Plots from rs6820223 Genotyping Assay ..... 24

Figure 5. Distribution of Library Complexity as a QQ-Plot ..... 25

Figure 6. Fragment Size Distribution of ATAC Libraries ..... 26

## Preface

I would like to thank previous and current members of the Martinson, Rappocciolo, Mailliard, and Rinaldo labs for their assistance throughout my time at the University of Pittsburgh, Diana DeLucia (Campbell), PhD, Patrick Mehta, B.S., Shivkumar Biradar, M.S., and Renee Anderko, B.S. I would like to specifically thank Shivkumar and Renee for their patience in teaching cell culture techniques and my advisor Dr. Martinson for his continued support. Finally, I would like to recognize the generous participants of the MACS cohort for their donations, without which, this research would not be completed.

# 1.0 Introduction

Human Immunodeficiency Virus (HIV), the causal agent of Acquired Immune Deficiency Syndrome (AIDS) was discovered in the early 1980s after five homosexual men from Los Angeles suffered from *Pneumocystis carinii* pneumonia (PCP) (1). This previously unknown illness was characterized by sexual transmission, opportunistic infections, and a rare type form of cancer known as Kaposi's sarcoma (2). By 1983, HIV was determined as the causative agent, and in 1984 the Multicenter AIDS Cohort Study was initiated with a goal of studying HIV pathogenesis as well as disease progression (3). Since the early years of HIV/AIDS, advances such as combined antiretroviral therapy (cART) and pre-exposure prophylaxis (PrEP) have greatly improve clinical outcomes for those infected with HIV-1 and those who are at high risk, however; a cure remains elusive (4, 5). Like many diseases, HIV-1 infection outcomes are influenced by the host response to infection, with some individuals possessing an inherent ability to control their infection in the absence of therapy (6). Studying these rare individuals may shed new light on HIV-1 therapy and treatment while simultaneously identifying novel mechanisms of viral control.

## 1.1 Natural History of HIV Infection

In 2014, the CDC revised the case definition for HIV-1 infection in the United States and divided the stages of disease into five categories based on the CD4+ T cell count. Stage 0 is defined by a negative HIV-1 test result within 6 months of acquiring a positive HIV-1 diagnosis. An individual is at Stage 1 when it has been 6 months since their positive HIV-1 diagnosis and CD4+

T cell counts are greater than or equal to 500 cells/μL of plasma. Stage 2 is defined by CD4+ T cell counts between 200-499/μL of plasma, and less than 200 cells/μL of plasma is considered Stage 3 which is often accompanied by an opportunistic infection due to immunodeficiency. The fourth stage of HIV-1 infection is "unknown" classified by an unknown CD4+ T cell count and T cell percentage in which case the stage of HIV-1 infection is undetermined (7). While disease progression and symptoms can vary, acute viremia at Stage 0 can result in flu-like symptoms including; headache, fever, and myalgia. Simultaneously, viral titers in the plasma increase, CD4+ T cells counts decline and a vigorous CD8+ T cell response develops leading to the resolution of acute infection symptoms. Following the acute phase of infection, the chronic phase begins about three to four months later resulting in suppressed viral titer, and a slow decline in CD4+ T cells eventually leading to the development of AIDS at Stage 3 (8).

## 1.2 HIV Pathogenesis

HIV-1 is often acquired via transmission at mucosal surfaces such as vaginal, rectal, penile, and rarely oral transmission (9-12). Once transmitted, the virus typically binds to cells expressing the primary receptor molecule CD4; primarily CD4+ T cells, macrophages and monocytes. Viral entry is achieved via the interaction of viral proteins gp120 and gp41 with the human secondary receptor proteins CCR5 and CXCR4 (13). Upon entry, the virus disassembles releasing two copies of the single-stranded RNA genome, and the viral proteins; reverse transcriptase (RT), integrase (IN) and protease (PR). Viral RNA becomes reverse transcribed into DNA and the viral IN enzyme incorporates the viral genome into the host genome (14, 15). Using host machinery, the viral genome gets transcribed into mRNA that are then translated into HIV-1 precursor proteins.

PR cleaves the polyprotein structures into individual viral proteins and the virus begins to assemble at the membrane surface of the host cell (16). After the viral proteins and RNA genome migrate to the newly formed particle, the virus buds from the cell in search of a new host cell to infect (17).

While CD4+ T cells are the primary target of HIV-1, CD4+ T cell depletion is primarily caused by programmed cell death of uninfected cells through multiple mechanisms. Monocytes undergoing CD4 cross-linking express Fas ligand resulting in CD4+ T cell apoptosis (18). Similarly, crosslinking of bound gp120 on the surface of uninfected CD4+ T cells induces cell death (19). Further, HIV-1 pathogenesis is driven by viral proteins including Nef. Nef is known to contribute to immune evasion by downregulating CD4 and MHC-1 molecules. In addition, Nef increases viral replication by activating NF-κB, JNK, and AP-1; transcription factors that bind to the HIV-1 long terminal repeat (LTR) leading to viral transcription (20). Finally, chronic immune activation contributes to HIV-1 pathogenicity. Chronic immune activation is believed to be due to multiple factors including the rapid depletion of CD4+ T cells in the gut disrupting the immunological barrier and allowing microbes and microbial products into the lymphatics; a phenomenon known as microbial translocation (21). Chronic immune activation may cause reduced CD4+ T cell-recovery on antiretrovirals and can contribute to early onset of non-AIDS related complications such as cardiovascular disease, and neurocognitive impairment (21-23).

### 1.3 Host Genetics and Progression to AIDS

Like many diseases, HIV-1 disease progression relies on viral pathogenicity as well as innate host genetics. Progression to AIDS can vary, and there are several host factors that contribute to rapid or slow progression. Many of these genetic factors interact directly with the

virus during its life cycle or are a part of the host immune response to the virus. For example, an important co-receptor for HIV-1, CCR5, is expressed on the surface of macrophages, dendritic cells, and CD4+ T cells and allows for attachment and entry. Individuals who are homozygous for the CCR5-Δ32 deletion (1% of the Caucasian population) demonstrate almost complete resistance to HIV-1 infection as they lack the co-receptor required for entry (24, 25). Those who are heterozygous for the deletion express lower levels of the CCR5 protein and have markedly slower progression to AIDS as the virus is unable to disseminate as effectively throughout the body (26).

As mentioned, variants in the immunological response may also dictate disease progression. Antigen presentation proteins called Human Leukocyte Antigen (HLA) are responsible for presentation of viral components thereby stimulating a specific and robust immune response. Variants in HLA-1 (A-C) receptors such as HLA*B57 are associated with slower progression. Individuals expressing HLA*B57 are more likely to present with asymptomatic acute HIV-1 infection and control HIV-1 viremia (<5000 copies HIV-1 RNA/mL) in the absence of therapy (27). However, variants and homozygosity in HLA types may result in rapid progression; such is the case for HLA*B35 (28).

### 1.4 Multicenter AIDS Cohort Study Participants

As previously mentioned, limited information about HIV-1 disease progression can be obtained from newly-infected individuals as, due to the advent of powerful and safe antiretroviral therapies (ARTs), patiens now commence therapy immediately upon diagnosis. Therefore, in this study we utilized retrospective patient samples from the Multicenter AIDS Cohort Study (MACS). The MACS is a 30 yearlong study of HIV-1 infection and disease progression in men who have

sex with men. Over 7,000 men have participated in the MACS sites of Baltimore, Chicago, Pittsburgh, and Los Angeles, and retrospective samples were used based on progression status (29). The participants utilized in this study include, Nonprogressors (NPs) ($n$ = 7) and Progressors (PRs) ($n$ = 6). NPs maintain low levels of viremia (<2,000 copies/mL) with stable CD4+ T cell counts (>500 cells/mm$^3$) over an average of 11 years after seroconversion. PRs had T cell counts <500 cells/mm$^3$ over approximately 5 years of infection with persistent viremia (30). NPs and PRs were ART naïve upon sample collection and were HLA typed and genotyped for CCR5Δ32. One NP but no PRs carried the CCR5Δ32 heterozygous genotype, and no NPs but two PRs, harbored the HLA B*5701 allele (30).

### 1.5 *Cis* vs. *Trans* Infection

HIV-1 disseminates from cell to cell via two different mechanisms. *Cis* infection is a replication dependent mechanism of transfer where virions produced from one host cell encounter another and initiate infection upon gp120/gp41 binding CD4 and CCR5 (31). On the contrary, *trans* infection involves the process of a cell mediated transfer of HIV-1 virions through an infectious synapse, and/or exocytosis of exomes carrying HIV-1, independent of viral replication (32). Of the antigen presenting cells (APCs) capable of *trans* infection, dendritic cells (DC) are particularly important because of their role linking the innate immune response to the adaptive by transfer of antigens to CD4+ helper T cells. Once a dendritic cell encounters HIV-1 in the periphery, HIV-1 binds to many C-type lectin receptors on the DC, such as DC-SIGN, via gp120 and the virus can either "surf" on the surface of the DC until it reaches a CD4+ T cell or become internalized and then released in an HIV-1 associated exosome (33).

Dendritic cells primarily facilitate *trans* infection as primary infection of DCs is severely limited *in vivo* (34). Dendritic cells, monocytes, macrophages, and resting CD4+ T cells express SAMHD1, an HIV-1 restriction factor. SAMHD1 is a deoxynucleotide triphosphate (dNTP) triphosphohydrolase responsible for limiting the pool of available dNTPs for viral replication (35). SAMHD1 and other restriction factors make *in vitro* infection of DCs limited and with low efficiency (36). Despite inefficient infection of DCs during *in vitro* studies, levels of plasmacytoid dendritic cells (pDCs) and myeloid dendritic cells (mDCs) correlate with disease progression following antiretroviral therapy (ART), suggesting a continuous disease process affecting DC viability regardless of clinical disease progression (37).

### 1.6 Epigenetic Regulation of Gene Expression

A multitude of factors contribute to gene expression and the process is highly complex. Among the mechanisms of gene expression, epigenetics refers to heritable variations that are not due to sequence level mutations (38). This is possible due to larger genome associated proteins called histones that are responsible for packaging DNA into tight bundles. Histones function as an octamer, with 2 copies of each histone protein: H2A, H2B, H3, and H4. An octamer of histones associated with DNA is referred to as a nucleosome. These epigenetic structures are important for gene regulation because they determine which regions of the genome are open and available to transcription enzymes such as transcription factors and RNA polymerase (39). Regulation of chromatin availability is an intricate process involving chromatin regulators which are further classified into histone modifying enzymes and chromatin remodeling proteins. Histone modifying enzymes are responsible for post translational modifications of histone tails via acetylation,

methylation, phosphorylation, and glycosylation, to name a few. While each modification works via a different mechanism, the end results is either open or close regions of chromatin in response to various stimuli. Chromatin remodeling proteins differ from histone modifying enzymes in that they primarily change the nucleosome structure by moving or removing histone proteins (39, 40). Chromatin remodeling proteins will not be further discussed as they are beyond the scope of this thesis work.

While there are many different methods for studying chromatin accessibility, Assay for Transposase-Accessible Chromatin (ATAC-Seq) is the only method able to provide a genome-wide analysis of the chromatin landscape (41). More established methods, including Chromatin Immunoprecipitation Sequencing (ChIP-Seq) involve antibody "pull-downs" which enrich for specific regions of the genome associated with particular histone modifications. While an effective technique, ChIP-seq cannot capture the entire genomic landscape (42).

ATAC-sequencing provides genome-wide chromatin accessibility data with as little as 50,000 cells--much less than is required for other genomic assays. The transposition reaction involves a modified bacterial enzyme known as transposase that carries sequencing adapters. During the reaction the enzyme will bind to open regions of chromatin and insert sequencing adapters required for downstream sequencing on Illumina platforms (41).

Because only sequences with adapters will be compatible with the downstream sequencing technology, other closed regions of genomic DNA will not be sequenced. Finally, data analysis involves a process known as "peak-calling" in which reads will be aligned to the reference genome and "peaks" will be called when reads consistently align at the same region of the genome (43). Peaks, therefore, represent regions that are open and available for transcription.

## 1.7 Genes of Interest

Rappocciolo et al. extracted total RNA from 9 NPs and 7 PRs and transcript levels were analyzed using an Illumina HT-12 BeadChip microarray. The Lumi and Limma software packages available in R were used to normalized and analyze the data resulting in the heat map in Figure 1. NPs (depicted here as VC, for Viremic Controller) are present underneath the red horizontal header and PRs are depicted underneath the blue horizontal header. Genes listed on the right side of the heat map were significantly differentially expressed in monocyte-derived DCs with an FDR of $q < 0.0025$. The heat map colors green and red indicate level of transcript expression with red indicating downregulation or a decrease in expression and green indicating higher levels of expression in a transcriptome-wide comparison of gene expression.

PALLD, also known as *Palladin*, was selected for SNP analysis as it showed the greatest difference in expression between NPs and PRs with higher expression in NPs, as seen in Figure 1.

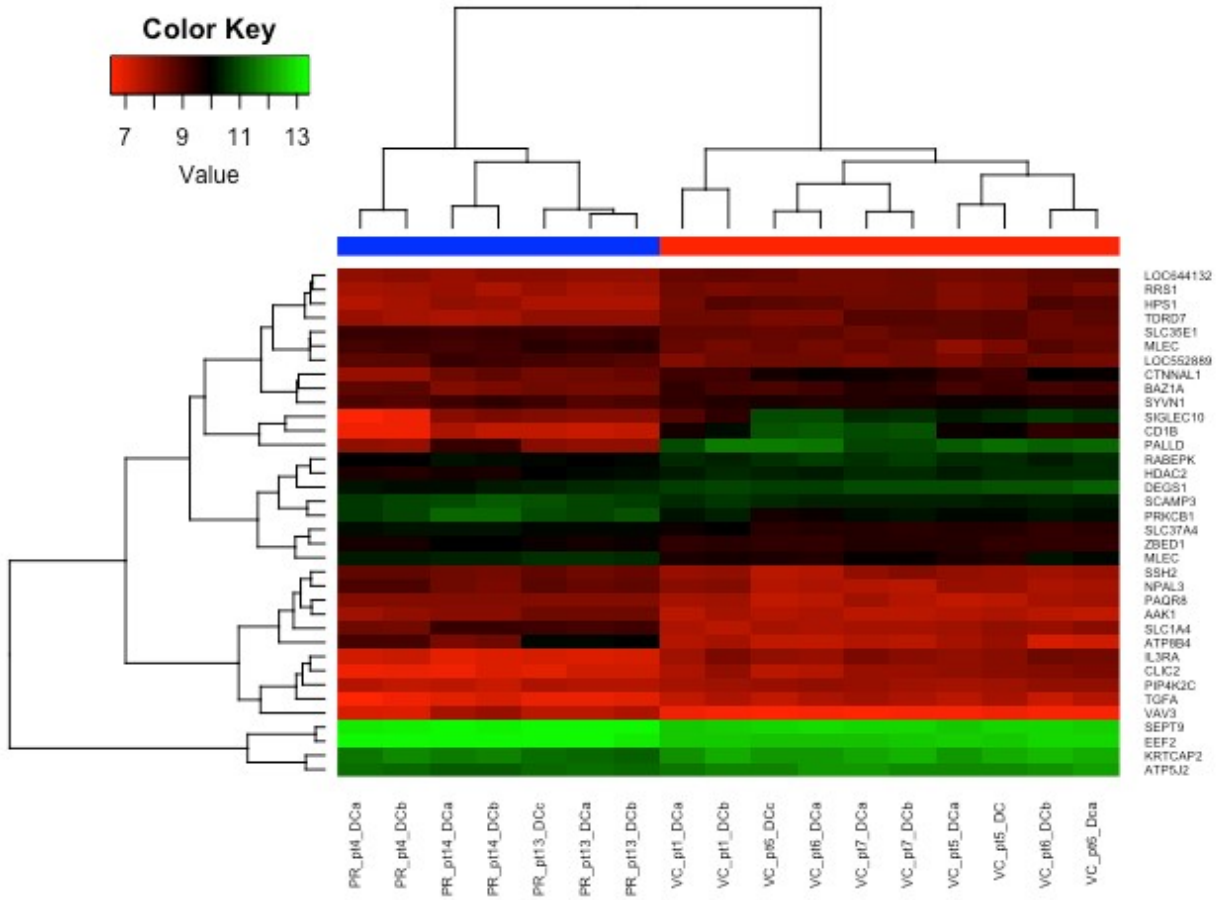


**Figure 1. Differentially expressed genes between MACS NPs and PRs.**

### 1.7.1 Palladin (PALLD)

Palladin is a recently characterized cytoskeletal protein involved in countless cytoskeletal rearrangements including, cell motility, filopodia formation, and membrane ruffling (44-46). The *PALLD* encodes 9 different isoforms with differential expression depending on cell state and type (47). The 90-Kda isoform, also known as isoform 4, binds filamentous actin using immunoglobulin domains at the C-terminal end, and binds actin-binding proteins VASP, α-actinin, and Eps8 at the N-terminal end (48). Isoform 4 is also the isoform known to be expressed in dendritic cells (49). Due to its ability to bind and recruit multiple actin-binding proteins, palladin is considered a

cytoskeletal scaffold protein conferring large effects on the cellular cytoskeleton when its overexpressed (50). The *PALLD* gene is about 431kb in length and is located on chromosome 4 with 3 transcriptional start sites (TSS) (48) .

## 2.0 Statement of Project

We hypothesize that differential gene expression observed between iDCs from NPs and PRs is due to promoter region mutations in *PALLD*. Further, we believe this differential gene regulation, and differences in disease progression may also be due in part to variants in chromatin regulators affecting chromatin structure, therefore; we are interested in using the Omni-ATAC-seq protocol for chromatin accessibility analysis. For many HIV NPs the mechanism behind disease control remains unknown. We aim to determine mechanisms behind differential gene regulation through sequencing level and epigenetic analyses.

### 2.1 Aim #1

Determine if there are promoter region mutations in *PALLD* that are associated with progression status.

### 2.2 Aim #2

Validate the use of the Omni-ATAC-seq protocol with healthy donor monocyte-derived iDCs.

# 3.0 Research Design

## 3.1 Amplicon Sequencing

Primer sequences to be used in amplicon sequencing of palladin, were generated utilizing the Primer3 program implemented in the commercial sequence analysis software, MacVector. Gene sequences were uploaded into MacVector from the University of California Santa Cruz Genome Browser. Separate primer sets were designed for each exon of *PALLD* as presented in Table 1. Oligonucleotide sequences were provided by Integrated DNA Technologies. *PALLD* amplicons were approximately 5Kb-3Kb in length to capture the promoter region including upstream regulatory regions of each transcriptional start site.

**Table 1. Primer Sequences for *palladin* Transcriptional Start Sites**

| GENE | REGION | SEQUENCE |
|---|---|---|
| PALLD | Transcriptional Start Site 1 | F: TTCAGGATCTCTGATTAACTACC |
| | | R: CAGATTAACAAACCAACACTAGC |
| | Transcriptional Start Site 2 | F: GTAAGGATCTGAAATGTGTCTTAG |
| | | R: ATAAATAGGGAGTACAGCTCTTCG |
| | Transcriptional Start Site 3 | F: CATTTTGAGCTAGATAGTGTCC |
| | | R: AGTTAAAATGAGACAAACGGG |

PCR conditions varied depending on amplicon length, and primer annealing temperatures. All PCR reactions were performed using Q5 High fidelity 2X Master Mix by New England Biolabs and samples were ran in triplicates. PCR conditions for each amplicon are represented in Table 2, Table 3, and Table 4. Components of each PCR reaction are found in Table 5.

**Table 2. PCR Conditions for Transcriptional Start Site 1 of *palladin***

| PCR Conditions | |
|---|---|
| 98°C | 30 secs |
| Then 35 cycles of: | |
| 98°C | 10 secs |
| 62°C | 30 secs |
| 72°C | 5.5 mins |
| Followed by: | |
| 72°C | 3 mins |

**Table 3. PCR Conditions for Transcriptional Start Site 2 of *palladin***

| PCR Conditions | |
|---|---|
| 98°C | 30 secs |
| Then 35 Cycles of: | |
| 98°C | 10 secs |
| 61°C | 30 secs |
| 72°C | 3.5 mins |
| Followed by | |
| 72°C | 3 mins |

**Table 4. PCR Conditions for Transcriptional Start Site 3 of *palladin***

| PCR Conditions | |
|---|---|
| 98°C | 30 secs |
| Then 35 Cycles of: | |
| 98°C | 10 secs |
| 61°C | 30 secs |
| 72°C | 5.5 mins |
| Followed by: | |
| 72°C | 3 mins |

**Table 5. *palladin* PCR Reaction Components**

| Reagent | Volume Per Reaction (µL) |
|---|---|
| Q5 High-Fidelity Buffer | 5 |
| 10 mM dNTPs | 0.5 |
| 10 mM Primers | 1.25 |
| Q5 High-Fidelity Polymerase | 0.25 |
| Q5 Enhancer | 5 |
| Nuclease Free Water | 11 |

Following PCR amplification, gel electrophoresis was performed on an 8 µL aliquot of the PCR reaction. The PCR reaction was mixed with 2 µL of loading dye and loaded into a 1% agarose gel with gel red. Electrophoresis was performed at 100 volts for 30 mins and fragments were visualized using a transilluminator. A representative gel electrophoresis image can be found in Figure 2.

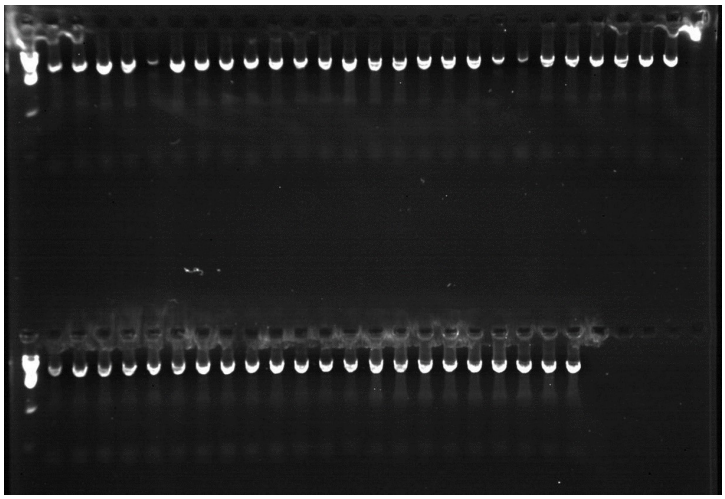

**Figure 2. *palladin* Transcriptional Start Site 1 Amplicons of 16 NPs, PRs and SN in Triplicate**

Amplicons were purified to remove PCR reaction contaminants using AMPure XP beads following the manufacturer's protocol. Briefly, libraries were suspended in 30.6 µL of AMPure

XP beads and incubated at room temperature for 5 mins. Libraries were placed on a magnetic rack for an addition 5 minutes at room temperature. Supernatant was removed, and bead pellets were washed twice with 70% ethanol while on the magnetic rack. Ethanol was removed, and beads were dried at room temperature for 3-5 minutes. Samples were removed from the magnetic rack and resuspended in 50 µL of 10 mM Tris HCl for 5 mins. Samples were placed back on the magnetic rack and supernatant was removed and stored at -20°C. Following purification, samples were quantified using Qubit Fluorometer 2.0 and amplicons from each transcriptional start site were pooled together resulting in one library of amplicons per individual. Library concentration and molarity prior to final pooling during library preparation can be found in Table 6. Molarity was calculated using the following equation based on average fragment size as determined by the Tapestation 2200 during library preparation:

$$\frac{\left(\frac{ng}{\mu L} X\ 10^6\right)}{\left(660\frac{g}{mol} X\ 600\ bps\right)}$$

Library preparation was performed using the Illumina Nextera Flex Kit for 24 samples following the manufacturer's protocol. The final library was quantified via Qubit Fluorometer prior to loading at 4nM with a 20% PhiX spike following the Denature and Dilute protocol for the MiSeq. Sequencing was performed on the Illumina MiSeq provided by the Human Genetics Department at the University of Pittsburgh's Graduate School of Public Health.

**Table 6. Library Concentration and Molarity for Amplicon Sequencing**

| Library | Concentration (ng/µL) | Molarity |
|---|---|---|
| PR 1 | 12.5 | 31.5 |
| PR 2 | 13.0 | 32.8 |
| PR 3 | 16.5 | 41.6 |
| PR 4 | 14.3 | 36.1 |
| PR 5 | 12.3 | 31.1 |
| NP 1 | 9.60 | 24.2 |
| NP 2 | 8.90 | 22.4 |
| NP 3 | 11.6 | 29.5 |
| NP 4 | 9.80 | 24.7 |
| NP 5 | 13.1 | 33.1 |
| NP 6 | 13.6 | 34.3 |
| NP 7 | 10.8 | 27.3 |
| SN 1 | 9.54 | 24.1 |
| SN 2 | 12.6 | 31.8 |
| SN 3 | 10.0 | 25.2 |

Data analysis was performed on Ubuntu 16.04.4 LTS. Alignment of the fastq files to the human hg38 reference genome was performed using BWA mem version 0.7.12-5. SAMtools version 0.1.19-1ubuntu1 was used to add coordinates to the mapped reads and index before GATK version 4.0.11.0 was used to mark duplicate reads, remove sequencing errors, call variants using Haplotype caller, and hard filter variants using parameters as follows: QD > 2, MQ < 40, FS > 60, Haploytype score > 13, MQRankSum < -12.5 and ReadPOSRankSum < -8. Joint variant calling files (VCF) were compared in Microsoft Excel and variants with distinguishing allele frequencies were determined.

## 3.2 TaqMan® Genotyping

Variants discovered in palladin were subjected to TaqMan based genotyping in a larger cohort of MACS participants to determine if allele frequencies observed in our small cohort may be replicated in a larger cohort. Using a TaqMan genotyping assay provided by ThermoFisher Scientific, 112 NPs and 58 PRs were genotyped following the manufacturer's protocol as described in Table 7. The genotyping assay was functionally tested and able to detect the A/G transition substitution variant with VIC (A-allele) and FAM (G- allele) dyes. Allelic discrimination plots were generated for each qPCR run (Figure 4). A contingency table was used to determine statistical significance when comparing the allele frequency in NP and PR cohorts (Table 9).

**Table 7. TaqMan® Genotyping Assay qPCR Components**

| Reagent | Volume Per Reaction (µL) |
|---|---|
| TaqMan Genotyping Master Mix | 12.5 |
| 20X Assay Working Stock | 1.25 |
| DNA Template | 2.5 |

**Table 8. TaqMan® Genotyping Assay qPCR Cycling Conditions**

| Cycling Conditions | |
|---|---|
| 95°C | 10 mins |
| The 40 cycles of: | |
| 95°C | 15 secs |
| 60°C | 1 min |

### 3.3 ATAC-Sequencing Cell Culture

Healthy donor monocytes were previously isolated using CD14+ selection from Central Blood Bank donor AP-001 are were used for ATAC-seq validation. Cells were removed from liquid nitrogen and thawed in a 37° water bath for 3 mins before being resuspended in 10 mL of complete IMDM media. Cells were pelleted at 1,200 rpm for 5 mins and resuspended in 5 mL of complete IMDM media for cell counting via hemocytometer. Half a million monocytes were then plated per well of a 12 well plate with IMDM media and 1,000 IU of IL-4 and GM-CSF for 5 days at 37°C. After 5 days iDCs were harvested by washing wells with 1X PBS and cells were counted via a hemocytometer.

### 3.4 Transposition Reaction

Triplicates were made of 50,000 cells per transposition reaction and iDCs were treated following the Omni-ATAC-seq protocol from Corces *et al*, 2017. Briefly, cells were incubated with 200U/mL of DNase I for 30 mins at 37°C then pelleted and washed with PBS to remove residual DNase. Cells were then resuspended in 50 µL of ATAC-Resuspension Buffer (RSB) containing 0.1% NP40, 0.1% Tween-20, and 0.01% Digitonin. After a 3 min incubation on ice, cells were resuspended in 1 mL of ATAC-RSB containing 0.1% Tween-20 and the tubes was inverted to mix. Nuclei were pelleted at 500 RCF for 10 mins at 4°C and ATAC-RSB was removed. Pellets were then resuspended in 50 µL of transposition mixture containing 2X TD buffer, transposase, PBS, 1% digitonin, 10% Tween-20, and nuclease free water. The reaction was incubated at 37°C for 30 mins in a thermomixer at 1000 RPM. DNA resulting from the

transposition reaction was purified using the Zymo DNA Clean and Concentrator-5 kit following the manufacturer protocol. DNA was stored at -20°C until amplification. An aliquot from each library was taken for fragment analysis via the Tapestation 2200. Concentrations from Tapestation 2200 analysis prior to library amplification are provided in Table 9.

**Table 9. Library Concentrations via Tapestation 2200**

| ATAC-seq Library | Concentration (pg/µL) |
|---|---|
| ATAC 1 | 21300 |
| ATAC 2 | 1150 |
| ATAC 3 | 765 |

### 3.5 ATAC Library Preparation and Sequencing

Libraries were amplified using 2X NEBNext Master Mix, as outlined in Table 10, and Illumina primer sequences with barcode indexes were used for multiplexing as seen in Table 11, Table 12, and Table 13. PCR conditions were as follows: 72°C for 5 mins, 98°C for 30 sec, then 5 cycles of 98°C for 10 sec, 63°C for 30 sec, and 72°C for 1 min. Reactions were then stored on ice while 5 µL was taken for qPCR amplification to determine the additional number of cycles required for complete library amplification. The 5µL aliquot of each library was amplified as follows: 98°C for 30 sec, then 35 cycles of 98°C for 10 sec, 63°C for 30 sec, and 72°C for 1 min. Reagents for the qPCR can be found in Table 7. Plotting RN vs. cycles, we determined which cycle yielded ¼ of total fluorescence intensity and amplified the libraries for an additional 11 cycles. Following amplification libraries were purified using Agencourt AMPure XP beads

following the size selection protocol by the Kaestner Lab at the University of Pennsylvania to remove fragments greater than 1000 bps.

**Table 10. Pre-Amplification PCR Master Mix**

| Reagent | Per Reaction (µL) |
|---|---|
| 25µM Primer i7 | 2.5 |
| 25µM Primer i5 | 2.5 |
| 2X NEBNext Master Mix | 25 |
| DNA Template | 20 |

**Table 11. Nextera Index Kit PCR Primers**

| | |
|---|---|
| **Index 1 Read** | CCAAGCAGAAGACGGCATACGAGAT**[i7]**GTCTCGTGGGCTCGG |
| **Index 2 Read** | AATGATACGGCGACCACCGAGATCTACAC**[i5]**TTCGTCGGCAGCGTC |

**Table 12. Nextera Index Kit Index 1 (i7) Adapter**

| i7 Index Name | Bases in Adapter |
|---|---|
| N701 | TCGCCTTA |

**Table 13. Nextera Index Kit Index 2 (i5) Adapters**

| i5 Index Name | Bases in Adapter |
|---|---|
| N502 | CTCTCTAT |
| N503 | TATCCTCT |
| N505 | GTAAGGAG |

**Table 14. Reagents for qPCR to Determine Additional Cycles**

| Reagent | Per Reaction (µL) |
|---|---|
| Primer i7 | 3.76 |
| Primer i5 | 0.5 |
| 25X SYBR Green | 0.5 |
| 2X NEBNext Master Mix | 0.24 |
| Water | 5 |
| DNA Template | 5 |

Libraries were sequenced on an Illumina NextSeq 550 sequencer, at the University of Pittsburgh Genomics Research Core, at concentration of 1.6pM using 2 x 76 cycles paired end with PhiX spiked in at 1%.

### 3.6 ATAC-Sequencing Data Analysis

Alignment of FastQ files was performed using the human reference genome hg19 and the latest version of BWA as of March 2018. SAMtools was used to fix and sort the genome based on gene coordinates and ATAC-seqQC version 3.8 in R was used for library coverage and quality analysis.

# 4.0 Results

## 4.1 rs6820223 in *palladin* is Present in All NPs and Absent in PRs

Palladin promoter region analysis revealed that SNP rs6820223 (A/G) located near the third transcriptional start site had an allele frequency of 0.929 in NPs but was completely absent from PRs. Allele frequency describes the number of chromosomes in the cohort that possessed the variant, meaning; 6 NPs were (AA) homozygous while 1 NP was (AG) heterozygous. A representative image from the Integrative Genomic Viewer (IGV) depicts homozygosity of rs 6820223 in NP 4 represented as an orange line throughout the read depth at position 168,832, 569 (Figure 3). rs6820223 is present in the RB1 transcription factor binding site near the isoform 4 promoter region of *palladin*; the known isoform to be expressed in iDCs (49).

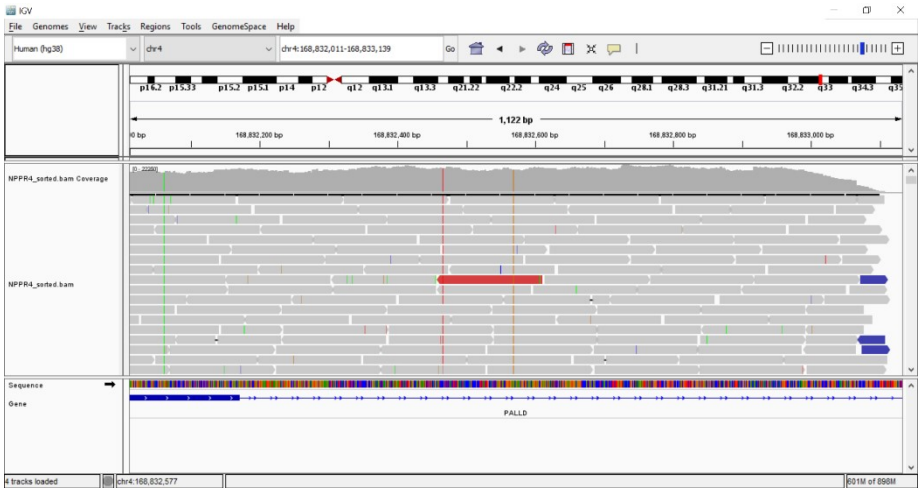


**Figure 3. Representative IGV Image of NP 4 Near Transcriptional Start Site 3**

## 4.2 *palladin* Variant is in the RB1 Transcription Factor Binding Site Near Isoform 4 Promoter Region

As mentioned, palladin possesses three alternative transcriptional start sites and expression is dependent on cell type and state. rs6820223 is located near the promoter region of the third transcriptional start site in the Retinoblastoma Protein 1 (RB1) binding site. RB1 is a tumor suppressor gene important for regulation of cell cycle and stabilization of heterochromatin.

## 4.3 rs6820223 Allele Frequency Was Not Replicated in a Larger MACS Cohort

Genotyping of SNP rs6820223 was performed using a commercially available TaqMan based genotyping kit provided by ThermoFisher Scientific. Using a probe-based design, the TaqMan assay determined the genotype at chromosome position 168,832,569 in a larger MACS cohort of 112 NPs and 58 PRs. Allelic discrimination plots were generated per qPCR run as seen in Figure 4. Odds ratios and p-values were calculated using a contingency table as seen in Table 15. Based from the contingency table, there is no statistically significant difference of the AA genotype between NPs and PRs (OD 0.96, p-value 0.94).

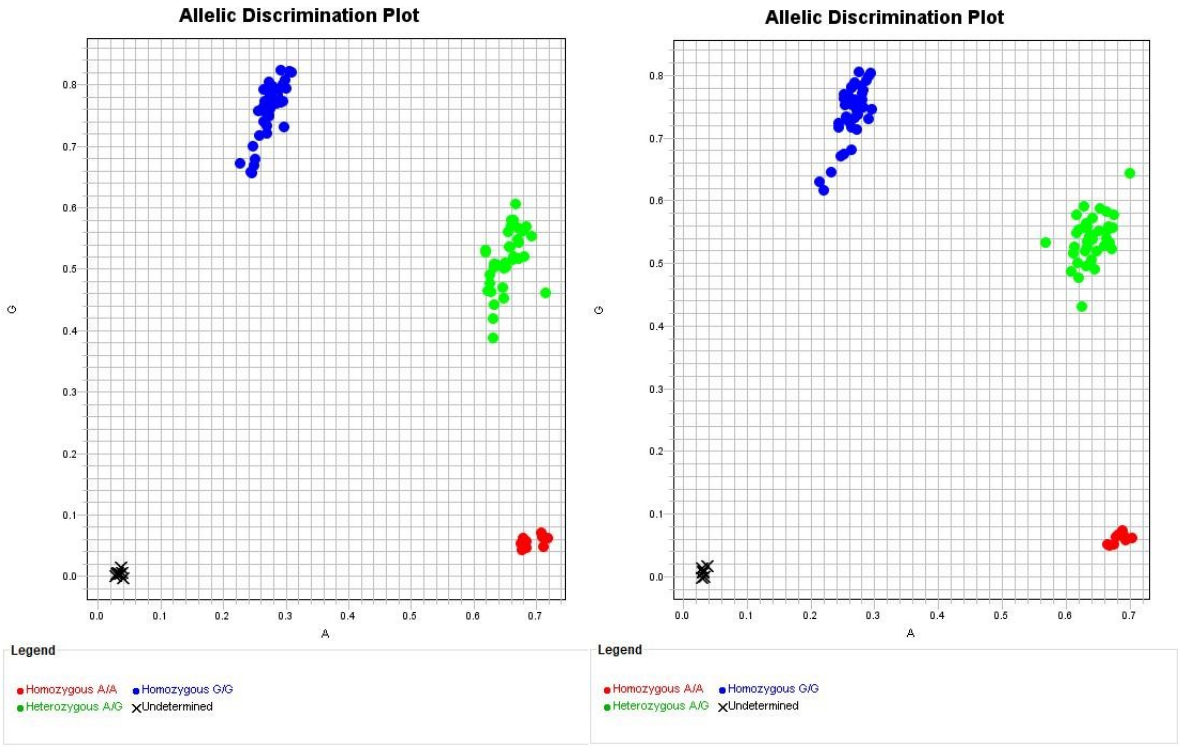


**Figure 4. Allelic Discrimination Plots from rs6820223 Genotyping Assay**

**Table 15. Allele Frequency Table of *palladin* rs6820223 in NPs and PRs**

| | NP | PR |
|---|---|---|
| **GG** | 54 | 26 |
| **GA** | 46 | 26 |
| **AA** | 12 | 6 |

## 4.4 ATAC-Sequencing Library Complexity and Mapping

ATAC-sequencing was performed on previously isolated monocytes from healthy donor AP-001 in triplicate. Libraries generated from iDCs had comparable quality metrics displayed in table 16. Percent proper pair describes the percentage of paired end reads that properly align, percent mapped describes the number of reads that were mappable to the human reference genome

hg19, mitochondrial rate describes the number of reads that mapped to the mitochondrial genome, and the PCR Bottleneck Coefficient conveys library complexity. A representative image of ATAC1 read fragment size is presented in Figure 6. Fragment sizes range from <100 bps to approximately 300bps with the large majority around 200 bps in length. As seen in the normalized read density plot, reads demonstrate periodicity with fragment sizes trailing after 300 bps.

**Table 16. Summary Statistics of the 3 ATAC-seq Libraries**

| | % Mapped | Mitochondrial Rate | Proper Pair Rate | PCR Bottleneck Coefficient |
|---|---|---|---|---|
| ATAC1 | 99.63 | 0.0098 | 0.965 | 3.87 |
| ATAC2 | 99.62 | 0.0250 | 0.966 | 2.90 |
| ATAC3 | 99.57 | 0.0446 | 0.974 | 2.85 |

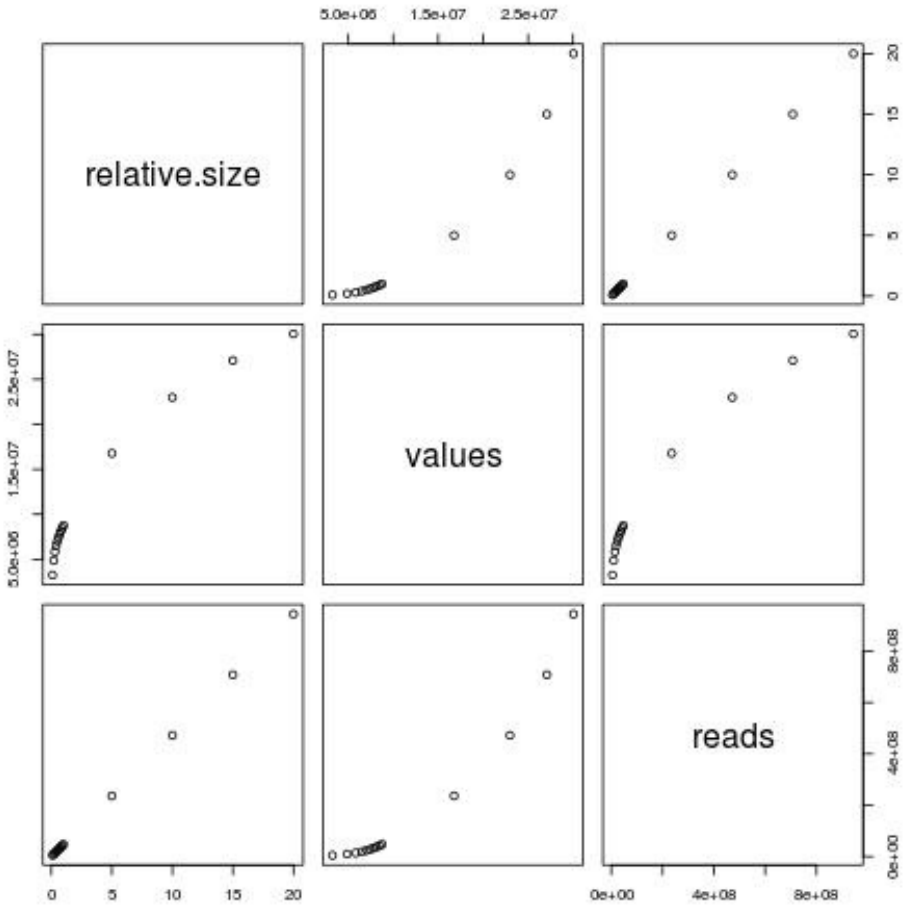


**Figure 5. Distribution of Library Complexity as a QQ-Plot**

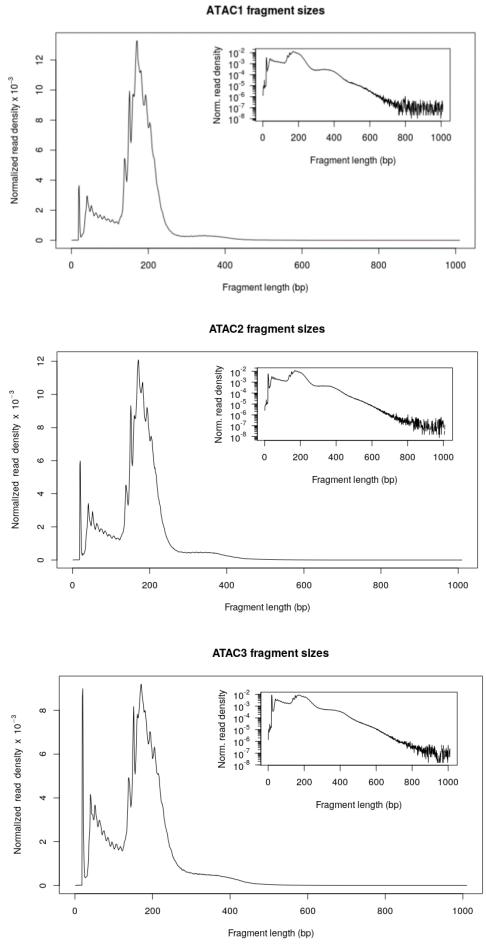


**Figure 6. Fragment Size Distribution of ATAC Libraries**

# 5.0 Discussion

Since the 1990s, HIV-1 treatment and outcomes have vastly improved with the advent of combined antiretroviral therapies. Yet, in the early years of HIV-1 infection it became clear that progression to AIDS was not uniform in the HIV-1 infected population. Individuals known as progressors represented the typical rate to AIDS progression while a smaller group now known as nonprogressors were able to maintain stable CD4+ T cell counts with varying levels of viral load in the absence of therapy. Recent studies have shown that a select group of MACS nonprogressors possess iDCs that are inherently unable to trans infect CD4+ T cells. Understanding underlying variants and epigenetic contributions to the HIV-1 response may uncover novel mechanisms of control to be applied to HIV-1 treatment, and possibly to viral infections broadly.

## 5.1 Transcriptome Analysis Determines Palladin Expression in Nonprogressors with Inefficient *Trans* Infection is Significantly Higher than in Progressors

Recent studies completed by Rappocciolo et. al determined that a select group of MACS cohort nonprogressors were unable to disseminate HIV-1 via *trans* infection of CD4+ T cells *in vitro*. These nonprogressors are not homozygous for the CCR5-Δ32 mutation, nor do they possess the HLA-B57 subtype. Because this select group of nonprogressors did not possess the known causes of non-progression, it is possible these individuals control their HIV-1 infection through novel mechanisms. HT-12 microarray data obtained by Rappocciolo et al, revealed that *palladin* expression was significantly upregulated in the nonprogressor cohort with a p-value $< 0.0025$ as

presented in Figure 1. It is important to note that due to the 3' IVT probe detection design of the HT-12 microarray, it was not able to discern which palladin isoform was overexpressed in the nonprogressors. Due to this limitation, amplicon sequencing was performed on each transcriptional start site to identify variants that may impact the numerous Palladin isoforms leading to overexpression.

### 5.1.1 Palladin is a Cytoskeletal Scaffold Protein with Drastic Effects on Cellular Function

Previous studies have described palladin as a cytoskeletal scaffold protein essential for numerous rearrangements including filopodia, membrane ruffles, and stress fiber formations depending on the cell type (48). Of the nine different isoforms it was determined that iDCs primarily expression isoform 4 which stems from the third transcriptional start site of *palladin (44)*. While many studies have conflicting conclusions on the primary role palladin plays, it is clear, among other cytoskeleton proteins that concentration within the cell is delicate and tightly controlled. Palladin overexpression and knockdown has significant impact on normal cellular functions. For example, overexpression of palladin within the context of cancer increases cellular motility, while palladin overexpression can also contribute to cell rigidity via the strengthening of stress fibers(47, 51). Interestingly, Zaccard et al. recently determined that tunneling nanotubes (TNTs) potentially act as an effective mechanism of *trans* infection from type-1 polarized dendritic cells to CD4+ T cells *in vitro* (52). Tunneling nanotubes are cytoskeletal projections that connect DCs to T cells upon CD40L stimulation (52). While palladin remains an understudied protein, it is feasible to consider the impact of *palladin* overexpression on the ability to form TNTs and effectively *trans* infect.

## 5.2 Amplicon Sequencing Detects Variant in the Third TSS of palladin with a High Allele Frequency in NPs

Amplicon sequencing of the three *PALLD* transcriptional start sites revealed variant rs6820223 near the third TSS of palladin with an allele frequency of 0.929 in the NP cohort, and 0 in the PR cohort. Allele frequency describes the number of chromosomes within the cohort that possess the variant; meaning all but one NP were homozygous for the variant and the variant was not present in the PR cohort. The variant is located in the binding site for RB1, a known transcription factor important for controlling heterochromatin structure for gene expression. As previously mentioned, iDCs primarily express isoform 4 which originates from the third TSS so it is possible rs6820223 disrupts RB1 binding leading to increased expression of isoform 4. Further studies into the impacts of chromatin structure on gene expression in HIV NP and PR cohorts will be needed to determine the functional impact of epigenetic influences on HIV-1 disease progression.

## 5.3 Allele Frequency of rs6820223 in Modest Cohort was not Replicated in a Larger Cohort Analysis

Genotyping of rs6820223 was performed in a larger MACS cohort consisting of 112 NPs and 58 PRs using a Taqman probe-based assay. Genotyping revealed that rs6820223 was not enriched in the NP cohort as seen in the small cohort that was used for amplicon sequencing. While it is possible rs6820223 is not associated with progression status, it is also possible that the cohort of 112 NPs possess other genetic causes of nonprogression. A heterogenous cohort of NPs can

reduce the signal of an association between rs6820223 and nonprogression. A larger cohort of NPs must be assayed for trans-presentation capability, *palladin* expression, cellular cholesterol content and, genotyped for rs6820223 to fully investigate the relationship. As for many genetic association studies, a genetically heterogenous population can be problematic for detecting true associations between genotype and phenotype therefore; researchers must make efforts to carefully classify and group cohorts based on phenotypes such as *palladin* expression, and cholesterol content.

### 5.4 The Omni-ATAC-Sequencing Method Successfully Removes Mitochondrial Contamination and can be Used to Determine Chromatin Accessibility in NPs and PRs

Assay for transposase accessible chromatin (ATAC) sequencing is a relatively new method to determine open regions of chromatin from cells or a cell line of choice. Similar to ChIP-sequencing, ATAC-seq results in sequencing reads aligning to a whole reference genome with sequencing depth determining regions called “peaks”. Peaks are regions of chromatin that were accessible by the transposase enzyme to adhere sequencing adapters and are therefore considered open and available for transcription. The Omni-ATAC-sequencing protocol published by Corces et al. provides a modified ATAC-seq protocol that may be used on a wide range of cell types and allows for improved signal-to-background ratio. The protocol includes cell/nucleus washes in Digitionin, Tween 20, and NP40 as well as a DNase step to remove DNA from non-viable cells.

The protocol was applied to iDCs differentiated from healthy donor monocytes to determine if the protocol could be used on MACS retrospective NP and PR samples. ATAC-sequencing was chosen over other chromatin assays because it allows for as little as 25,000 cells

as input and can reveal the genome-wide chromatin landscape as opposed to ChIP-sequencing which requires specific regions to be "pulled-down".

The Omni-ATAC-seq protocol was followed as stated in Corces et al, utilizing the DNase step to improve signal-to-background ratio, and Illumina Nextera XT adapters were used for library preparation as stated in Table 11, Table 12, and Table 13. Following sequencing from the Genomic Core Labs at the University of Pittsburgh, reads were aligned to the hg19 reference genome using BWA and sorted and indexed using SAMTools before being loaded into ATACseqQC. ATACseqQC is an R/Bioconductor program used to assess library complexity and quality from ATAC-sequencing protocols. Proper library complexity and quality indicates a successful sequencing experiment. Because our data were obtained from triplicate preparations from one donor, we do not expect to see differences between them, other than what might be due to technical variation in sample processing.

The ATACseqQC program was used to generate summary statistics for each library as provided in Table 16. Of importance, mitochondrial reads were vastly depleted using the Omni-ATAC-seq protocol in all libraries. Similarly, the percent of reads mapped was significantly high for each library indicating a successful sequencing run. The PCR bottleneck coefficient determines how skewed the data is based on read counts across the genome, and while the distribution is slightly skewed from normal as seen in Figure 5, sequencing depth was sufficient across the genome. Further, the ENCODE ATAC-sequencing data quality standards indicate that each library should contain approximately 50-million reads for paired end sequencing and 2 out of the 3 libraries had greater than 60-million reads, while ATAC library 1 was comparable with 47-million reads. Another important parameter for ATAC-sequencing data is fragment size distribution as shown in Figure 6. Fragment sizes for a successful ATAC-seq run are typically between 50-1,000

bps with a decline in density as the fragments get larger. Per Corces et al, 2018, ATAC-seq libraries with a fragment distribution similar to Figure 6 are typically displaying biased size selection during library preparation. Moving forward, the AMPure Bead size selection protocol should be modified to remove bias and improve the concentration of reads less than 100 bps and greater than 300 bps.

## 5.5 Experimental Limitations

As with many experiments, the most relevant cell types are not always available due to cell availability limitations. Monocyte-derived immature dendritic cells behave similarly to *in vivo* immature dendritic cells however, it is important to note that the majority of *in vivo* immature dendritic cells are not monocyte-derived. Therefore, inconsistencies may exist between our *in vitro* results and what may be occurring *in vivo*. Immature dendritic cells were chosen for study to be consistent with the HT-12 microarray data from Figure 1 which was completed on MACS iDCs.

## 5.6 Conclusion and Future Directions

Understanding how innate host genetics contributes to disease progression is important for a plethora of infectious diseases. More importantly, genetic characteristics associated with slower or faster disease progression may not always lie in immunological pathways. Discovery tools such as transcriptome analysis, and chromatin accessibility performed on small, and highly characterized cohorts may reveal nonintuitive mechanisms of disease control in need of further investigation. *Palladin* is a novel gene in HIV progression studies and overexpression of palladin

in our cohort of NPs warrants an investigation of palladin expression in a larger cohort. Such NPs with high palladin expression may then be genotyped for rs6820223 to investigate an association. Additionally, given palladin's role in cytoskeletal rearrangements, immunohistochemistry of palladin expression during iDC differentiation and activation may inform its role in general function and TNT development. Secondly, ATAC-seq performed in MACS NP and PR iDCs will reveal inherent differences in chromatin structure that may present new avenues of investigation.

## Bibliography


1. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (CDC). 1981. Pneumocystis pneumonia--Los Angeles. MMWR Morb Mortal Wkly Rep 30:250-2.
2. Beral V, Peterman TA, Berkelman RL, Jaffe HW. 1990. Kaposi's sarcoma among persons with AIDS: a sexually transmitted infection? Lancet 335:123-8.
3. Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rozenbaum W, Montagnier L. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 220:868-71.
4. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, Sullivan AK, Clarke A, Reeves I, Schembri G, Mackie N, Bowman C, Lacey CJ, Apea V, Brady M, Fox J, Taylor S, Antonucci S, Khoo SH, Rooney J, Nardone A, Fisher M, McOwan A, Phillips AN, Johnson AM, Gazzard B, Gill ON. 2016. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. Lancet 387:53-60.
5. Autran B, Carcelain G, Li TS, Blanc C, Mathez D, Tubiana R, Katlama C, Debre P, Leibowitch J. 1997. Positive effects of combined antiretroviral therapy on CD4+ T cell homeostasis and function in advanced HIV disease. Science 277:112-6.
6. Gonzalez E, Dhanda R, Bamshad M, Mummidi S, Geevarghese R, Catano G, Anderson SA, Walter EA, Stephan KT, Hammer MF, Mangano A, Sen L, Clark RA, Ahuja SS, Dolan MJ, Ahuja SK. 2001. Global survey of genetic variation in CCR5, RANTES, and MIP-1alpha: impact on the epidemiology of the HIV-1 pandemic. Proc Natl Acad Sci U S A 98:5199-204.
7. U. S. Department of Health and Human Services, Centers for Disease Conrol Prevention (CDC). 2014. Revised surveillance case definition for HIV infection--United States, 2014. MMWR Recomm Rep 63:1-10.
8. Clark SJ, Saag MS, Decker WD, Campbell-Hill S, Roberson JL, Veldkamp PJ, Kappes JC, Hahn BH, Shaw GM. 1991. High titers of cytopathic virus in plasma of patients with symptomatic primary HIV-1 infection. N Engl J Med 324:954-60.
9. Edwards S, Carne C. 1998. Oral sex and the transmission of viral STIs. Sex Transm Infect 74:6-10.
10. Patterson BK, Landay A, Siegel JN, Flener Z, Pessis D, Chaviano A, Bailey RC. 2002. Susceptibility to human immunodeficiency virus-1 infection of human foreskin and cervical tissue grown in explant culture. Am J Pathol 161:867-73.
11. Rerks-Ngarm S, Pitisuttithum P, Ganguly N, Zhang L, Tamashiro H, Cooper DA, Vun MC, Bela B, Ditangco R, Van Kinh N, Bernstein A, Osmanov S, Mathieson B, Kent SJ, Shao Y. 2010. Defining the objectives of the AIDS vaccine for Asia network: report of the WHO-UNAIDS/Global HIV vaccine enterprise regional consultation on expanding AIDS vaccine research and development capacity in Asia. Curr Opin HIV AIDS 5:435-52.
12. Figdor CG, van Kooyk Y, Adema GJ. 2002. C-type lectin receptors on dendritic cells and Langerhans cells. Nat Rev Immunol 2:77-84.
13. Zaitseva M, Blauvelt A, Lee S, Lapham CK, Klaus-Kovtun V, Mostowski H, Manischewitz J, Golding H. 1997. Expression and function of CCR5 and CXCR4 on human Langerhans cells and macrophages: implications for HIV primary infection. Nat Med 3:1369-75.
14. Varmus H. 1987. Reverse transcription. Sci Am 257:56-9, 62-4.

15. Imamichi T. 2004. Action of anti-HIV drugs and resistance: reverse transcriptase inhibitors and protease inhibitors. Curr Pharm Des 10:4039-53.
16. Brik A, Wong CH. 2003. HIV-1 protease: mechanism and drug discovery. Org Biomol Chem 1:5-14.
17. Karacostas V, Wolffe EJ, Nagashima K, Gonda MA, Moss B. 1993. Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. Virology 193:661-71.
18. Oyaizu N, Adachi Y, Hashimoto F, McCloskey TW, Hosaka N, Kayagaki N, Yagita H, Pahwa S. 1997. Monocytes express Fas ligand upon CD4 cross-linking and induce CD4+ T cells apoptosis: a possible mechanism of bystander cell death in HIV infection. J Immunol 158:2456-63.
19. Banda NK, Bernier J, Kurahara DK, Kurrle R, Haigwood N, Sekaly RP, Finkel TH. 1992. Crosslinking CD4 by human immunodeficiency virus gp120 primes T cells for activation-induced apoptosis. J Exp Med 176:1099-106.
20. Varin A, Manna SK, Quivy V, Decrion AZ, Van Lint C, Herbein G, Aggarwal BB. 2003. Exogenous Nef protein activates NF-kappa B, AP-1, and c-Jun N-terminal kinase and stimulates HIV transcription in promonocytic cells. Role in AIDS pathogenesis. J Biol Chem 278:2219-27.
21. Sankaran S, George MD, Reay E, Guadalupe M, Flamm J, Prindiville T, Dandekar S. 2008. Rapid onset of intestinal epithelial barrier dysfunction in primary human immunodeficiency virus infection is driven by an imbalance between immune response and mucosal repair and regeneration. J Virol 82:538-45.
22. Kamat A, Lyons JL, Misra V, Uno H, Morgello S, Singer EJ, Gabuzda D. 2012. Monocyte activation markers in cerebrospinal fluid associated with impaired neurocognitive testing in advanced HIV infection. J Acquir Immune Defic Syndr 60:234-43.
23. Deeks SG. 2011. HIV infection, inflammation, immunosenescence, and aging. Annu Rev Med 62:141-55.
24. Rana S, Besson G, Cook DG, Rucker J, Smyth RJ, Yi Y, Turner JD, Guo HH, Du JG, Peiper SC, Lavi E, Samson M, Libert F, Liesnard C, Vassart G, Doms RW, Parmentier M, Collman RG. 1997. Role of CCR5 in infection of primary macrophages and lymphocytes by macrophage-tropic strains of human immunodeficiency virus: resistance to patient-derived and prototype isolates resulting from the delta ccr5 mutation. J Virol 71:3219-27.
25. Michael NL, Louie LG, Rohrbaugh AL, Schultz KA, Dayhoff DE, Wang CE, Sheppard HW. 1997. The role of CCR5 and CCR2 polymorphisms in HIV-1 transmission and disease progression. Nat Med 3:1160-2.
26. Eugen-Olsen J, Iversen AK, Garred P, Koppelhus U, Pedersen C, Benfield TL, Sorensen AM, Katzenstein T, Dickmeiss E, Gerstoft J, Skinhoj P, Svejgaard A, Nielsen JO, Hofmann B. 1997. Heterozygosity for a deletion in the CKR-5 gene leads to prolonged AIDS-free survival and slower CD4 T-cell decline in a cohort of HIV-seropositive individuals. AIDS 11:305-10.
27. Altfeld M, Addo MM, Rosenberg ES, Hecht FM, Lee PK, Vogel M, Yu XG, Draenert R, Johnston MN, Strick D, Allen TM, Feeney ME, Kahn JO, Sekaly RP, Levy JA, Rockstroh JK, Goulder PJ, Walker BD. 2003. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. AIDS 17:2581-91.
28. Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, Kaslow R, Buchbinder S, Hoots K, O'Brien SJ. 1999. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. Science 283:1748-52.
29. Detels R, Jacobson L, Margolick J, Martinez-Maza O, Munoz A, Phair J, Rinaldo C, Wolinsky S. 2012. The multicenter AIDS Cohort Study, 1983 to. Public Health 126:196-198.
30. Rappocciolo G, Jais M, Piazza P, Reinhart TA, Berendam SJ, Garcia-Exposito L, Gupta P, Rinaldo CR. 2014. Alterations in cholesterol metabolism restrict HIV-1 trans infection in nonprogressors. MBio 5:e01031-13.

31. Turville SG, Santos JJ, Frank I, Cameron PU, Wilkinson J, Miranda-Saksena M, Dable J, Stossel H, Romani N, Piatak M, Jr., Lifson JD, Pope M, Cunningham AL. 2004. Immunodeficiency virus uptake, turnover, and 2-phase transfer in human dendritic cells. Blood 103:2170-9.
32. Chen Q, Swaminathan S, Yang D, Dai L, Sui H, Yang J, Hornung RL, Wang Y, Huang da W, Hu X, Lempicki RA, Imamichi T. 2013. Interleukin-27 is a potent inhibitor of cis HIV-1 replication in monocyte-derived dendritic cells via a type I interferon-independent pathway. PLoS One 8:e59194.
33. Bracq L, Xie M, Benichou S, Bouchet J. 2018. Mechanisms for Cell-to-Cell Transmission of HIV-1. Front Immunol 9:260.
34. Nobile C, Petit C, Moris A, Skrabal K, Abastado JP, Mammano F, Schwartz O. 2005. Covert human immunodeficiency virus replication in dendritic cells and in DC-SIGN-expressing cells promotes long-term transmission to lymphocytes. J Virol 79:5386-99.
35. St Gelais C, de Silva S, Amie SM, Coleman CM, Hoy H, Hollenbaugh JA, Kim B, Wu L. 2012. SAMHD1 restricts HIV-1 infection in dendritic cells (DCs) by dNTP depletion, but its expression in DCs and primary CD4+ T-lymphocytes cannot be upregulated by interferons. Retrovirology 9:105.
36. Smed-Sorensen A, Lore K, Vasudevan J, Louder MK, Andersson J, Mascola JR, Spetz AL, Koup RA. 2005. Differential susceptibility to human immunodeficiency virus type 1 infection of myeloid and plasmacytoid dendritic cells. J Virol 79:8861-9.
37. Fontaine J, Coutlee F, Tremblay C, Routy JP, Poudrier J, Roger M, Montreal Primary HIVI, Long-Term Nonprogressor Study G. 2009. HIV infection affects blood myeloid dendritic cells after successful therapy and despite nonprogressing clinical disease. J Infect Dis 199:1007-18.
38. Handy DE, Castro R, Loscalzo J. 2011. Epigenetic modifications: basic mechanisms and role in cardiovascular disease. Circulation 123:2145-56.
39. Zhang P, Torres K, Liu X, Liu CG, Pollock RE. 2016. An Overview of Chromatin-Regulating Proteins in Cells. Curr Protein Pept Sci 17:401-10.
40. Mirabella AC, Foster BM, Bartke T. 2016. Chromatin deregulation in disease. Chromosoma 125:75-93.
41. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B, Kathiria A, Cho SW, Mumbach MR, Carter AC, Kasowski M, Orloff LA, Risca VI, Kundaje A, Khavari PA, Montine TJ, Greenleaf WJ, Chang HY. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14:959-962.
42. Angelini C, Costa V. 2014. Understanding gene regulatory mechanisms by integrating ChIP-seq and RNA-seq data: statistical solutions to biological problems. Front Cell Dev Biol 2:51.
43. Ou J, Liu H, Yu J, Kelliher MA, Castilla LH, Lawson ND, Zhu LJ. 2018. ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. BMC Genomics 19:169.
44. Parast MM, Otey CA. 2000. Characterization of palladin, a novel protein localized to stress fibers and cell adhesions. J Cell Biol 150:643-56.
45. Doughman RL, Firestone AJ, Wojtasiak ML, Bunce MW, Anderson RA. 2003. Membrane ruffling requires coordination between type Ialpha phosphatidylinositol phosphate kinase and Rac signaling. J Biol Chem 278:23036-45.
46. Goicoechea SM, Arneman D, Otey CA. 2008. The role of palladin in actin organization and cell motility. Eur J Cell Biol 87:517-25.
47. Goicoechea SM, Bednarski B, Stack C, Cowan DW, Volmar K, Thorne L, Cukierman E, Rustgi AK, Brentnall T, Hwang RF, McCulloch CA, Yeh JJ, Bentrem DJ, Hochwald SN, Hingorani SR, Kim HJ, Otey CA. 2010. Isoform-specific upregulation of palladin in human and murine pancreas tumors. PLoS One 5:e10347.

48. Dixon RD, Arneman DK, Rachlin AS, Sundaresan NR, Costello MJ, Campbell SL, Otey CA. 2008. Palladin is an actin cross-linking protein that uses immunoglobulin-like domains to bind filamentous actin. J Biol Chem 283:6222-31.
49. Mykkanen OM, Gronholm M, Ronty M, Lalowski M, Salmikangas P, Suila H, Carpen O. 2001. Characterization of human palladin, a microfilament-associated protein. Mol Biol Cell 12:3060-73.
50. Rachlin AS, Otey CA. 2006. Identification of palladin isoforms and characterization of an isoform-specific interaction between Lasp-1 and palladin. J Cell Sci 119:995-1004.
51. Cannon AR, Owen MK, Guerrero MS, Kerber ML, Goicoechea SM, Hemstreet KC, Klazynski B, Hollyfield J, Chang EH, Hwang RF, Otey CA, Kim HJ. 2015. Palladin expression is a conserved characteristic of the desmoplastic tumor microenvironment and contributes to altered gene expression. Cytoskeleton (Hoboken) 72:402-11.
52. Zaccard CR, Watkins SC, Kalinski P, Fecek RJ, Yates AL, Salter RD, Ayyavoo V, Rinaldo CR, Mailliard RB. 2015. CD40L induces functional tunneling nanotube networks exclusively in dendritic cells programmed by mediators of type 1 immunity. J Immunol 194:1047-56.