DISCOVERY AND CHARACTERIZATION OF ESSENTIAL VIRAL AND HOST PROTEINS FOR THE ONCOLOYTIC SENECA VALLEY VIRUS (SVV)

By

Linde A. Miles

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, MD

June 2016
Abstract

Seneca Valley Virus (SVV-001) is a newly discovered oncolytic virus that has the ability to selectively infect and kill cancer cells with neuroendocrine features, including small cell lung cancer (SCLC). SVV-001 has shown potential as an oncolytic therapy in in vitro assays, preclinical mouse models, and early phase clinical trials in both SCLC and pediatric neuroendocrine cancers. Currently, there are certain limitations to SVV-001 virotherapy that have delayed the progression of SVV-001 in the clinic. The research presented below describes two studies performed in the hopes of overcoming current obstacles for SVV-001 virotherapy progression.

The first study describes the characterization of substrate specificity of a SVV-001 encoded protease for the design of a protease-activated peptide prodrug to use in combination with SVV-001 virotherapy. Kinetic studies with both purified recombinant SVV-001 protease and SVV-001 infected SCLC cell lines established the protease substrate specificity and also determined the optimized protease substrate for the SVV-001 protease. The study presented in the second and third chapters describe the initial discovery and characterization of the cellular receptor for SVV-001. Two genome-wide loss-of-function CRISPR screens were performed and identified the anthrax toxin receptor 1 (ANTXR1) as an essential gene for SVV permissivity. Further loss-of-function and gain-of-function studies in both SCLC and pediatric cell lines confirmed that ANTXR1 is an essential protein for SVV permissivity and the major binding determinant of SVV on intact cells. Moreover, binding and co-immunoprecipitation studies as well as the resolution of a cryo-electron microscopy (cryo-EM) structure established that SVV interacts directly with ANTXR1. We have, for the first
time, identified the cellular receptor for SVV-001. Research presented in these studies can ultimately be used to further progress SVV-001 in the clinic as a treatment for SCLC and pediatric neuroendocrine cancers.

Advisor: Charles M. Rudin, M.D., Ph.D.

Thesis Reader: Samuel R. Denmeade, M.D.
Acknowledgments

I’d like to thank, first and foremost, my advisor and mentor Dr. Charlie Rudin for allowing me to complete my graduate thesis research in his lab. I am the scientist I am today because of the training I gained in your lab and the opportunities I had while in your lab. Thank you for your guidance over the past 6 years, for allowing me to be independent, for being a voice of positivity during some of the hardest days, and for always giving me your unending support and time. Thank you for allowing me to join your lab and also for inviting me to be a founding member of your lab at Memorial Sloan Kettering in New York. Being at MSK gave me so many opportunities and experiences and made my thesis project progress in ways I didn’t think was possible. I am extremely appreciative for everything you have done and continue to do for my project and for my scientific career. Thank you so much!

I also would like to thank my other mentor, Dr. J.T. Poirier. From the day we met in Charlie’s lab I was continuously impressed by your knowledge and creativity. Continuing the SVV project you started as your thesis project left very big shoes for me to fill. I hope I have done it justice. Thank you for always making time to discuss my project and for being a cheerleader and voice of reason when my project hit standstills. The day we FINALLY identified the receptor will always be one of my fondest memories!

I’d also like to thank my thesis committee: Dr. Sam Denmeade, Dr. Caren Meyers, and Dr. Michelle Rudek. Your guidance, input, and suggestions for my thesis project were so incredibly helpful over the past 6 years. I can’t thank you enough for everything you’ve contributed to my thesis project! Additionally, thank you to all of my collaborators and the principal investigators that have contributed to my thesis project and scientific career: Dr. Mihnea Bostina, Dr. Laura Burga, Dr. Nate Brennan, Dr. Megan Matthews, Dr. Ralph Garippa, and Dr. Ouathek Ouerefelli.

Having good labmates makes the years for graduate school go by a little easier and faster. Thank you to all past Rudin lab members who helped welcome me into the lab at Hopkins and quickly became not only colleagues, but also good friends. I especially need to thank my good friend and labmate both at Hopkins and MSK, Eric Gardner. The lab transition from Hopkins to MSK would have been so much harder had I not had a colleague and friend like you in lab with me every day. Thank you for helping me with mouse work, for always challenging me with new scientific questions, but most importantly for being a great friend! Thank you also to present Rudin lab members for helpful discussions regarding my project and career.

Thank you to my classmates from Hopkins for making our first years of grad school memorable and fun. I wish you all the best! Thank you also to the Pharmacology department support staff and MSK support staff for all of your help during my time in graduate school. Thank you to the many friends, both from pre-grad school and during grad school for your unending support (and patience) over the past 6 years. Having good friends makes life worth living-thank you all from the bottom of my heart.
Finally, I would not be here today if it wasn’t for the amazing encouragement, support, and cheerleading from my family especially my Mom, Dad, and sister Kelsey. From day one you have all pushed me to be a hard worker, to always challenge myself, and to be curious and use that curiosity to search for answers. I thank my mom and dad for getting me interested in science in the first place. Mom, Dad: You have been amazing role models and I can’t thank you enough for sharing in this experience with me and being there for me during both the happy and not so happy days. Kels: Thank you for always being there for me during the past 6 years. Every day I am thankful for you. I love all of you so much!
## Table of Contents

Abstract .................................................................................................................................................. ii

Acknowledgements ................................................................................................................................. iv

List of Tables .......................................................................................................................................... viii

List of Figures ........................................................................................................................................ ix

Introduction ........................................................................................................................................... 1

  Small Cell Lung Cancer (SCLC) ................................................................................................. 1

  Oncolytic Viruses ......................................................................................................................... 3

  Seneca Valley Virus (SVV-001) ............................................................................................... 8

Chapter 1. Seneca Valley Virus 3C<sup>pro</sup> Substrate Optimization Yields Efficient Substrates for use in Peptide-Prodrug Therapy ......................................................... 14

  Processing of Picornavirus Polyproteins .............................................................................. 14

  Genome and Polyprotein of SVV ......................................................................................... 17

  Protease Activated Peptide-Prodrugs ............................................................................... 19

  Study Introduction .................................................................................................................. 21

  Materials and Methods ......................................................................................................... 24

  Results ............................................................................................................................................. 36

  Discussion ....................................................................................................................................... 55

Chapter 2. Design, execution, and analysis of pooled <i>in vitro</i> CRISPR/Cas9 screens: Identifying essential genes for SVV ...................................................... 60

  History of Genome Wide Screens ...................................................................................... 60

  The CRISPR/Cas System .................................................................................................. 62

  Types of Human Genome Wide CRISPR Libraries .......................................................... 64

  Choice of Cell Line for Pooled CRISPR Screening .......................................................... 72
List of Tables

Table 1. Oligonucleotide list for primers used for the construction of FRET fusion proteins and recombinant SVV-001 3Cpro .............................................................27

Table 2. Summary of second order rate constants ($k_{cat}/K_M$) for the two endogenous substrates, controls and L/VP4 amino acid substitution/truncation mutants ........39

Table 3. Oligonucleotide list for primers used in cloning and sequencing for CRISPR constructs and ANTXR1 expression constructs ...........................................89
List of Figures

Fig. 1. Conceptual schematic of use of an SVV 3C^pro activated peptide prodrug in combination with SVV virotherapy as a novel form of VDEPT ..........................22

Fig. 2. SDS-PAGE gel of overexpression and purification of recombinant fusion SVV-001 3C^pro and catalytically dead mutant SVV-001 C160A 3C^pro ..................30

Fig. 3. SVV-001 polyprotein and identification of efficient 3C protease substrates ..........38

Fig. 4. Conversion of Substrates after Incubation with Increased Concentration of SVV-001 3C^pro .................................................................40

Fig. 5 Conversion of Substrates after Incubation with SVV-001 3C^pro compared to SVV-001 C160A 3C^pro .................................................................42

Fig. 6 Comparison of SVV 3C^pro- L/VP4.1 Substrate Specificity to Selected Picornavirus Protease-Substrate Pairs ..............................................44

Fig. 7 Comparison of Conversion of Substrate containing Q↓SP cleavage site to L/VP4 and L/VP4.1 Substrates ..........................................................45

Fig. 8. Optimization of the L/VP4 Substrate ..............................................47

Fig. 9. SVV-001 3C^pro substrates are cleaved in the context of a cellular infection ........50

Fig. 10. Stability of L/VP4.1 fluorescently quenched peptide in human plasma ..........54

Fig. 11. Types of CRISPR screens ..............................................................68

Fig. 12. Example data analysis of a genome-wide pooled CRISPR screen ..................78

Fig. 13. HAP1 cells are permissive to SVV ...............................................97

Fig. 14. Depiction of genome wide CRISPR knockout (GeCKO) screen workflow ........98

Fig. 15. Relative representation of each sgRNA in the Human GeCKO v2 library at various points during the screens ....................................................100

Fig. 16. Identification of ANTXRI as an essential host determinant for SVV ............103

Fig. 17. Knockout of the ANTXRI gene leads to the loss of SVV permissivity ..........124

Fig. 18. ANTXRI expression is significantly associated with permissivity ............127

Fig. 19. Activity of IFN response pathways can alter SVV permissivity in
Introduction

Small Cell Lung Cancer (SCLC)

Lung cancer is the leading cause of cancer related mortalities annually in the United States. There are two major subtypes of lung cancer categorized historically by their morphological appearance and biological differences: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). SCLC diagnoses account for approximately 15% of all lung cancer cases annually in the United States (1). Because of the highly aggressive nature of SCLC, the majority of patients present with extensive stage SCLC (ES-SCLC) and have local and/or distal metastases (2). The most common metastatic sites for SCLC are the lymph nodes, bones, and brain (1).

Standard line treatment for ES-SCLC is a platinum-based chemotherapy, either cisplatin or carboplatin, combined with a topoisomerase inhibitor, etoposide that causes DNA damage in tumor cells. The platinum based drugs intercalate within the DNA of cells causing crosslinks that also lead to severe DNA damage (3). Although the initial response to first-line treatment is high (approximately 60-70%), almost every patient relapses eventually with tumors that have acquired resistance to the platinum-etoposide chemotherapy. After relapse, the only Food and Drug Administration (FDA) approved second line treatment for SCLC is topotecan, a topoisomerase inhibitor similar to etoposide. Unfortunately, only 30% of patients respond to topotecan treatment, leaving many patients with few alternative treatment options (2, 4). Furthermore, over thirty late phase clinical trials in over forty years have failed to identify superior and more effective treatments for SCLC (5).
Compared to other cancers, SCLC is genetically characterized by a high mutation load usually due to tobacco exposure (6). Almost universal inactivation of two important tumor suppressors, tumor protein 53 (TP53) and retinoblastoma 1 (Rb1), is observed in SCLC (7-9). Additionally, v-myc avian myelocytomatosis viral oncogene homolog (MYC) family members oncogenes are altered or amplified in roughly 20% of SCLC tumors (10, 11). In contrast to NSCLC where the identification of distinct driver mutations have allowed for the stratification of patients for targeted therapy treatments, no targetable driver mutations are currently known for SCLC. Recent large scale genomic studies of SCLC cell lines, patient derived xenografts (PDX) and patient samples have identified genetic and epigenetic alterations that could lead to possible new targets for SCLC treatment (7-9, 12).

SCLC is considered a neuroendocrine cancer with relatively high expression of classical neuroendocrine markers, such as chromogranin A and synaptophysin, and morphologic similarities to cells of neuroendocrine origin, such as pulmonary neuroendocrine cells in the lung (4, 11, 13, 14). There are two histologic subtypes of SCLC, classic and variant, which have shown dependency on distinct transcriptional programs (15, 16). Two mutually exclusive basic helix-loop-helix (bHLH) transcription factors, ASCL1 and NEUROD1, have defined the classic and variant subtypes of SCLC (16-18). In certain instances, the subtypes show discrete differences to certain chemotherapies. Variant SCLCs are characterized by relatively low expression of ASCL1, increased expression of genes involved in neuronal differentiation, and more apt to be resistant to standard chemotherapies (12, 15, 16, 19). We have additionally shown that Seneca Valley Virus, a recently discovered oncolytic virus that infects and lyses cancer cells, has selective tropism for variant SCLC (20).
**Oncolytic Viruses**

Oncolytic viruses are defined as replication competent wild-type or genetically engineered viruses that selectively infect and kill cancer cells while sparing normal cells. These viruses can use multiple methods to kill and lyse infected tumor cells only after the virus has replicated and repackaged its genome. Indirect damage to nearby, uninfected tumor cells can also occur during infection. Viruses, in general, are very selective for which cells they infect based on a unique cellular tropism. Because of the selectivity, the majority of these oncolytic viruses are well tolerated by patients even at high doses, as much of the non-specific damage to normal cells seen with cytotoxic chemotherapies is not an issue (21).

Although the existence of viruses had been known since the beginning of the 19th century, it was not until the mid-1900s that viruses were being considered, and subsequently tested, as a possible cancer treatment. Much of the initial evidence that viruses may have the capability to infect and kill cancer cells came from anecdotal clinical responses and temporary remissions of cancer patients after certain viral infections, like influenza and hepatitis (22). In 1949, the first clinical trial examining the treatment of cancer by a virus was performed in Hodgkin’s lymphoma patients. Patients were infected with Hepatitis B virus after two previous patients underwent temporary remissions following a hepatitis infection. A minority of patients on study did observe improvements in certain disease symptoms and tumor shrinkage; however, most patients contracted hepatitis and more than one death due to viral infection occurred during the study (23). Other early clinical trials assessing the efficacy of viruses as cancer therapies were executed with similar conflicting results (24-26).
Around the same time as the early clinical trials, preclinical mouse models of cancers were being developed (27-29). Led by Alice Moore, multiple human and animal viruses with oncolytic potential were assayed in vivo (30-32). The use of mouse models also allowed the optimization of viral delivery method to the tumors. The ability of these early preclinical studies in identifying viruses with highest potential led to the standard “proof of principle” step of testing oncolytic viruses in in vivo cancer models in rodents prior to testing in patients (33, 34).

Just as there are many types of pathogenic viruses, there are also multiple types of oncolytic viruses that can be classified by genome or whether the viral genome has been genetically modified. Although early clinical trials were done using wild-type strains of viruses, many oncolytic agents being tested in preclinical and early phase clinical trials recently have been genetically altered in some way. The first generation of optimized viruses were wild-type strains adapted to infection of cancer cells by serial passaging. Based on assumptions by Moore and others, serial passaging of certain viruses allowed for mutational selection of clones with increased tumor specificity and lytic capacity (32-35). As recombinant DNA techniques became widely available, researchers attempted to further optimize oncolytic viruses by removing or attenuating virulence factors, directing tropism to specific cell types, and/or adding immune stimulatory cytokines to activate the immune system against the tumor (36, 37). Genetic manipulation of viral genomes can increase the complexity of approval procedures by the Food and Drug Administration (FDA), but can augment unwanted side effects and increase the efficacy of the recombinant virus (38). In fact, the first oncolytic virus to be approved by the FDA was talimogene laherparepvec (T-VEC), a
recombinant Herpes simplex virus (HSV) armed with the granulocyte-macrophage colony stimulating factor (GM-CSF) (39, 40).

Although many oncolytic viruses have shown promising efficacy as single agents, the production of neutralizing antibodies limits multiple dosing regimens over prolonged periods of time (41, 42). Recently, studies have shown that combining the lytic capacity of oncolytic viruses with various immunotherapies increase anti-tumor responses (43, 44). As the oncolytic virus lysed the infected tumor cells, tumor antigens are released into the tumor microenvironment and nearby circulation. The newly released tumor antigens can alert and activate both the innate and adaptive immune system in two major ways (39, 45). Firstly, the tumor associated antigens activate and direct cytotoxic T cells and macrophages to circumvent the immune suppressive environment within the tumor and directly remove tumor cells. Secondly, the tumor antigens activate the systemic immune system to target tumor cells at distal locations from the tumor (39). This phenomenon is known as an “immune associated” bystander effect of the virus infection (46). Additionally, the release of viral proteins from lysed cells also alert the immune system to an active virus infection, which elicits a strong reaction from immune cells (39, 41, 47). Due to the promising responses being observed in immunotherapy clinical trials, many oncolytic viruses are now being examined as combination treatments with immunotherapy agents (NCT01740297, NCT02043665).

The first and most common class of oncolytic viruses are the DNA viruses, which includes adenoviruses and herpesviruses. The viral genomes of these viruses are single or double stranded DNA, which are transcribed and translated into viral proteins. The genomes are also replicated using a DNA-dependent DNA polymerase, usually in the nucleus of the host cells.
As the wild-type versions of most of the oncolytic DNA viruses are known human pathogens, the vast majority of the oncolytic DNA viruses in clinical development have been genetically modified or attenuated (37, 49). DNA viruses have relatively large genomes, which allow for relatively easy genetic alteration or transgene addition (39). Genetic modifications can redefine tropism of the virus, increase replication and oncolytic spread of the virus, and/or decrease virulence and safety concerns of the virus (38, 50). Although these viruses can be easily modified, due to the nature of the viral genomes, unwanted insertional mutagenesis of the host genome can be a safety concern for some of the DNA viruses. Additionally, there is a higher chance of pre-existing antibodies to certain DNA viruses as many are common human pathogens. This can limit the efficacy of the virus or increase the potential of unwanted side effects as the antibodies may rapidly recognize and clear the virus from circulation upon administration (39).

The second and more varied class of oncolytic viruses are the RNA viruses, which contain an RNA molecule as genetic material. The oncolytic RNA viruses span multiple virus families that include poliovirus, reovirus, and Newcastle disease virus. Overall, the RNA viruses are smaller in size with correspondingly smaller genome sizes than the DNA viruses (48). The smaller genome size usually leads to more difficult introduction of transgenes into the genome, however, multiple researchers have been able to genetically alter the genomes of RNA viruses by deletion or alteration of virulence genes (51-53). For example, researchers have modified the poliovirus to decrease neurovirulence and increase tropism for glioblastoma cancer cells by utilizing an attenuated strain and replacing a section of the genome known as the internal ribosome entry site (IRES) with the IRES from a related human rhinovirus (HRV) (51, 54, 55). This recombinant virus, known as PVS-RIPO, is
currently being tested in a phase I clinical trial (NCT01491893). An additional advantage of RNA viruses as oncolytic therapies is the method of genome replication. RNA viruses utilize direct RNA replication in the cytoplasm for genome replication, which removes the potential for unwanted DNA integration into the host genome and possible insertional mutagenesis (48).

The RNA virus family, *picornaviridae*, contains multiple oncolytic viruses that have shown promising efficacy as cancer treatments. As described above, the genetically modified poliovirus is under investigation in glioblastoma (NCT01491893) (55, 56). Another oncolytic picornavirus currently in clinical trials is the wild-type coxsackievirus, specifically the A21 strain (CVA21) under the commercial name CAVATAK. One of very few unmodified viruses being tested, CAVATAK has a natural tropism for cancer cells as some cancer types overexpress the receptor for the virus (57, 58). Promising results have been observed in phase I and II clinical trials in melanoma (NCT00438009, NCT01227551). Early phase clinical trials testing the efficacy of CAVATAK have also begun in solid tumors (NCT02043665). The third oncolytic picornavirus of note is the native Seneca Valley Virus, which is the focus of the research described here.
**Seneca Valley Virus (SVV-001)**

Seneca Valley Virus (SVV-001), a newly characterized native picornavirus, has been shown to selectively infect and lyse cancer cells with neuroendocrine features, including SCLC and a range of pediatric brain tumors \((59, 60)\). SVV-001 was originally discovered as an accidental contaminant possibly introduced via porcine trypsin into cell culture. Alignment with other viruses found it most similar to the *picornaviridae* family, of which it is now a member \((59, 61)\). The Picornavirus family is comprised of non-enveloped, single stranded, positive RNA viruses that include the human rhinoviruses (HRV), cardioviruses, and poliovirus. As mentioned previously, two other picornaviruses are currently being tested as oncolytic viruses: Cavatak, the WT A21 strain of coxsackievirus, and PVS-RIPO, the genetically modified Sabin strain of poliovirus.

SVV-001 was initially tested in a range of normal and cancer cell lines and found to have high selectivity for approximately 50% of SCLC and pediatric neuroendocrine cancer cell lines, such as neuroblastoma and Ewing sarcoma \((59, 60)\). Permissive cells showed a range of susceptibility to SVV-001 with 50% effective concentrations (EC\(_{50}\)) from \(10^{-4}\) to \(10^{4}\) viral particles per cell. Additionally, no normal human cells were found to be permissive to SVV-001 \((59)\). Further studies demonstrated that SVV could successfully replicate exclusively in the cytoplasm of permissive cell lines and this lytic life cycle of SVV-001 was approximately 6-8 hr \((59, 61, 62)\).

Showing promising selectivity for cancer cell lines *in vitro*, SVV-001 was extensively tested in a range of preclinical mouse models of permissive cancer subtypes. In SCLC, initial studies were performed in athymic mice engrafted subcutaneously with the most permissive
cells line, the SCLC cell line H446. Mice bearing tumors were challenged with a single intravenous dose of SVV-001 ranging from $10^7$ to $10^{13}$ viral particles (vp) per kg. All mice receiving SVV-001 at a dose of $10^8$ vp/kg or higher showed complete H446 tumor eradication. Lower SVV-001 doses showed significant anti-tumor efficacy with complete regressions in a majority of H446 tumor bearing mice (59). Further studies in patient derived xenografts (PDX) models of SCLC demonstrated the specificity of SVV-001 for the variant subtype of SCLC. SVV-001 demonstrated significant anti-tumor efficacy in multiple permissive SCLC PDX models after a single intraperitoneal injection of $10^{12}$ vp/kg. Moreover, higher doses of SVV-001 caused durable tumor growth inhibition in all SVV-001 treated mice and complete tumor regression in over 30% of mice (20). Using an infectious, replication competent, GFP reporter clone of SVV-001 (SVV-GFP), the exquisite selectivity of SVV was demonstrated in H446 xenografts. GFP positive cells were only observed within the H446 tumor and not within other organs, such as the liver (62).

The efficacy of SVV-001 was additionally examined in multiple preclinical mouse models of pediatric cancer subtypes. Initially tested in the highly permissive retinoblastoma cell line Y79, complete regressions were observed in the majority of tumor bearing mice following a single doses of SVV-001 of $10^8$ vp/kg or higher (59). Anti-tumor efficacy was also confirmed in an orthotopic model of invasive retinoblastoma using immunocompetent Rag2 mice (63). Further in vivo studies were performed in the Pediatric Preclinical Testing Program (PPTP) subcutaneous xenograft panel. Permissive xenografts, the majority found to be tumor subtypes with neuroendocrine features, were found to have significantly delayed tumor growth after a single intravenous dose of SVV-001 ($3 \times 10^{12}$ vp/kg) (59, 60). In orthotopic permissive medulloblastoma models, SVV-001 dosed intravenously was found to be capable
of crossing the blood-brain barrier (BBB) and infecting medulloblastoma tumor cells. Permissive tumors infected with a single dose of SVV-001 significantly prolonged survival of tumor bearing mice and caused complete tumor regression in the majority of treated mice. Furthermore, the SVV-001 infection was confined to medulloblastoma tumor cells and did not expand to adjacent brain tissue (64). Overall, SVV-001 showed promising efficacy in multiple preclinical mouse models, indicating the potential of SVV as a virotherapy and helped to move this agent into clinical trials.

As the safety and efficacy of many oncolytic viruses depend on prior exposure of patients to the virus, randomly collected human blood samples were tested for the presence of neutralizing antibodies, indicating a prior exposure to SVV-001. Only 1 out of over 100 serum samples showed low levels of neutralizing antibodies, confirming the prevalence of prior SVV-001 exposures in humans is relatively rare (59). Studies also demonstrated that SVV-001 was not inactivated in human blood and incubation of the virus did not lead to hemagglutination. SVV-001 has been shown to be endemic in the pig population in the United States, but has been shown to not cause disease in pigs, mice, or humans (59). Additionally, a toxicology study was performed in both male and female A/J mice prior to SVV-001 being tested in early phase clinical trials. Mice were injected with a single dose of SVV-001 at multiple viral titers and evaluated for up to 12 weeks after administration of the virus. The only significant difference observed between vehicle and SVV-001 treated mice during the study was a dose dependent decrease of white blood cells in SVV-001 treated mice 24 hours after administration. The drop in white blood cells resolved without assistance by the 1-week post administration evaluation (59).
A phase I clinical trial further assayed the safety of SVV-001 in patients with advanced solid neuroendocrine tumors. Increasing single doses of SVV-001 were administered intravenously to patients. The study concluded that there were no dose limiting toxicities (DLT) during treatment of SVV-001, even in the cohorts given the highest doses of SVV-001. Two SCLC patients died as a result of progressive metastatic disease while on study. Autopsy concluded the deaths were not a result of viral infection. Furthermore, evaluation of one of these patients showed productive SVV-001 infections in extensive liver metastases 28 days post administration. However, no evidence of virus was observed in adjacent normal liver tissue or other cancer-free organs like the kidney and pancreas, confirming the ability of SVV-001 to home specifically to tumors through the vasculature. Although SCLC patients were only treated with the lowest dose of virus (10^7 vp/kg), analysis of serum samples of treated patients demonstrated prolonged production of the virus up to 3 weeks post SVV-001 administration. Moreover, serum results demonstrated increased viral titers up to 10,000-fold higher than the viral titer initially administered to patients in the SCLC cohort, assumedly due to successful infection of tumors and intratumoral replication of the virus. Patients developed neutralizing antibodies to SVV-001 after initial administration, which may limit the dosing strategies available for SVV-001. Serum virus titers correlated with a few cases of disease stabilization, small tumor reductions, and symptom improvements and resolutions confirming clinical activity of SVV-001 as a virotherapy (42).

A phase I clinical trial in children with relapsed or refractory neuroendocrine tumors followed the adult phase I clinical trial examining the safety of SVV-001 as a single agent (Part A) or in combination with cyclophosphamide (Part B). Part A was designed as a dose escalation study testing intravenous doses from 10^9 to 10^{12} vp/kg to determine the MTD in
children. The dose of SVV-001 identified as safe and tolerable in Part A (10^{11} \text{ vp/kg}) was utilized in Part B of the clinical trial. Part B combined two intravenous doses with SVV-001 separated by 21 days with two rounds of 14-day oral cyclophosphamide. SVV-001 as a single agent or given in combination was again found to be safe and tolerable. Cyclophosphamide did not significantly delay the productive of neutralizing antibodies in patients treated in Part B. Sustained viral replication was not observed past 15 days post administration. Although patients were not followed long term as in the adult phase I trial, stable disease was observed in 6 of 12 Part A and 4 of 6 Part B patients who completed the study (65). Overall SVV-001 was determined to be safe and tolerable in children with neuroendocrine tumors.

Multiple studies demonstrated the potential of SVV-001 as a safe and effective virotherapy in both preclinical mouse models and early phase clinical trials. Specific limitations have continued to hinder the clinical development of SVV-001 in both SCLC and pediatric patients. The first major restriction of SVV-001 is the generation of neutralizing antibodies that develop after a single administration of the virus. These antibodies bind and deactivate SVV-001 while helping to clear the virus from the circulation system (42). Administration of SVV-001 after the generation of neutralizing antibodies can also lead to more severe side effects. Therefore, repeated dosing of SVV-001 may be not be possible. The second and most important limitation of SVV-001 is the unknown method of tropism. Currently, it is not possible to identify which patients are permissive to an SVV-001 replication and therefore benefit from SVV-001 as a therapy (59, 60). Only recently was SVV permissivity in variant SCLC correlated with high \textit{NEUROD1} and low \textit{ASCL1} gene expression (20). However, essential proteins required for an SVV infection are still unknown. The studies described in
the following chapters detail the progress made to overcome each of the limitations for SVV-001 in the hopes of progressing clinical development of SVV-001 as a virotherapy.
Chapter 1. Seneca Valley Virus 3C\textsuperscript{pro} Substrate Optimization Yields Efficient Substrates for use in Peptide-Prodrug Therapy

Processing of Picornavirus Polyproteins

The Picornavirus family of viruses is comprised of a vast array of pathogenic agents that affect both humans and animals. Notable picornaviruses include the cause of the common cold-human rhinovirus, poliovirus, and foot-and-mouth disease virus (FMDV), a highly virulent virus that infects livestock animals. Picornaviruses, whose name means small RNA viruses, house their RNA genomes within protein capsids made of 60 subunits of four proteins (66-69). The viral genome of each of these viruses encodes a single polyprotein, which contains both structural and non-capsid proteins (67). The genomes are translated directly into protein using internal ribosome entry sites (IRES) found within the 5’ noncoding region of the RNA genome. IRES sequences have the capability to recruit ribosomes and protein involved in translation initiation in the absence of a 5’ cap, which is utilized for canonical translation initiation by host mRNAs (70-72).

Conserved among picornaviruses, the viral polyprotein is organized into 4 distinct sequential groups: L-P1-P2-P3 (73). The first protein found in some but not all picornavirus polyproteins is denoted the leader or L protein (61, 67). The P1 polypeptide is comprised of the four structural proteins, VP1, VP2, VP3, and VP4, that make up the mature capsid subunits. Based on multiple crystal structures, VP1-3 form a protomer that is organized on the solvent exposed side of the capsid with VP4 proteins interacting with the trimeric subunit on the internal side of the viral capsid (66, 68, 74). The function of the P2 group of viral proteins, comprised of the 2A, 2B, and 2C proteins, is poorly understood compared the other viral protein subgroups (67). The 2A peptide encodes a viral protease that assists in
polyprotein processing and host protein processing in some picornaviruses (75, 76). However, some picornaviruses like Seneca Valley Virus contain an extremely short 2A peptide with no protease activity (61). The P3 peptide encodes 4 proteins 3A, 3B, 3C, and 3D, that function in genome replication and virus maturation. The 3D protein serves as the RNA dependent RNA polymerase for picornavirus genome replication (77-79). The 3C peptide encodes a cysteine protease that is responsible for the majority of polyprotein processing steps during translation (80, 81). This protease will be discussed in further detail in a later section. The 3B protein functions as an RNA binding protein, classically named VPg, that interacts with the 5' of the viral genome (82). Proteolytic processing of the polyprotein occurs during translation, which allows for released proteins to perform their roles as soon as they are translated (83, 84).

The 3C protease (3C^pro), as mentioned previously, is responsible for almost all proteolytic cleavage events within the polyprotein maturation. Sequence analysis and crystal structure resolution of 3C proteases from multiple picornaviruses identified a protease structure similar to chymotrypsin, a classic serine protease (85-87). Serine proteases contain a conserved Ser-His-Asp catalytic triad of amino acids within the active site. Interestingly, the picornaviral 3C proteases also contain a universally conserved catalytic triad, but the serine residue is replaced with a cysteine residue creating a Cys-His-Asp/Glu catalytic triad (88, 89). Mechanistically, the cysteine is responsible for the nucleophilic attack of carbonyl moiety in peptide bond being cleaved. However, the exact proteolytic mechanism of picornaviral 3C proteases is still not fully understood (90). As many picornaviruses are pathogenic in nature, the 3C^pro, an extremely important enzyme for virus production, is an attractive target for anti-
viral drug design. Studies involving the substrate specificity of numerous 3C proteases have been performed in the hopes of identifying potential peptide based inhibitors (91, 92).

Genome sequence analysis, N-terminal sequencing of viral proteins, and in vitro assays with purified viral proteases, were historically used to identify putative protease substrates for various picornaviral 3C proteases (90, 93-96). Studies identified the recognition site within the 3C protease spans at least 8 residues on the polyprotein, four amino acids on each side of the cleavage site, denoted P4-P4’ (90). Moreover, these studies confirmed a preferred 3C<sup>pro</sup> cleavage site between glutamine-glycine (Q-G) amino acid pairs that was originally identified upon analysis of the first completed poliovirus sequence (97, 98). The Q-G cleavage motif, however, is not highly conserved among picornaviruses and the substrate constraints can vary widely between closely related viruses (67). For instance, while the poliovirus and HRV 3C proteases cleave almost exclusively between Q-G residues, the 3C protease of FMDV has a less constrained substrate specificity and can cleave Q-G, Q-L, Q-T, Q-I, and even E-G amino acid pairs (90, 94, 98). Overall, substrate specificity appears to be similar among closely related viruses of the same genus.
**Genome and Polyprotein of SVV**

The genome of SVV encodes a polyprotein of 2181 amino acids in length with RNA sequence and polyprotein organization most closely resembling those of the cardioviruses (61). The 5’ untranslated region (UTR) contains an IRES similar to that of hepatitis C virus and pestiviruses, which allows for translation of the genome in cytoplasm using host machinery (61, 99). Amino acid sequencing of the 3 main capsid proteins, VP1-3, demonstrated similar sizes and sequences to other related picornaviruses (61). Additionally, resolution of the crystal structure of SVV-001 confirmed tertiary structure similarity to the cardioviruses, but deviations in the loops of the capsid proteins unique to SVV-001 were also observed (61, 100, 101). Although the capsid proteins were directly sequenced, the sequences of additional proteins in the polyprotein were projected based on alignment with closely related picornaviruses (61). Where certain picornaviruses polyproteins contain two proteases, the 2A and 3C proteases, that assist in the release and maturation of the viral proteins, the 2A sequence of SVV-001 is only 9 amino acids long and most likely has no proteolytic activity. Moreover, the 2A-2B junction of the polyprotein contains an amino acid sequence conserved in picornaviruses that uses a ribosome skipping mechanism to release the P1-2A section of the polyprotein co-translationally from the 2B-2C-P3 section (61, 67, 102). Therefore, the only putative protease in the SVV-001 polyprotein based on sequence alignment is the 3C\textsuperscript{pro} (61). Unlike other picornavirus 3C proteases, the polyprotein substrates determined by sequence alignment have not been confirmed nor has substrate specificity of the SVV-001 3C\textsuperscript{pro} been investigated. The first study described in this chapter examined the substrate specificity of the SVV-001 3C\textsuperscript{pro} to identify an optimized substrate. Using the optimized
substrate for the viral $3C^{\text{pro}}$, we sought to design a protease activated peptide-prodrug that could be activated by the $3C^{\text{pro}}$ as a potential combination therapy with SVV virotherapy.
Protease Activated Peptide-Prodrugs

The concept of a prodrug, or a drug that has to be activated by some method within the body or a cell, was first described in 1958 by Dr. Adrien Albert (103). There are multiple types of prodrugs, but all are designed to improve either safety, specificity, or biodistribution issues of the parent drug (104). A peptide prodrug contains a peptide substrate moiety linked to a cytotoxic agent in a way that inactivates the drug. Only after selective cleavage of the peptide moiety by the target protease does the cytotoxic agent become active. Most protease activated peptide prodrugs utilize intrinsic human cellular proteases found preferentially in the target cell type versus all others (105).

Historically, the vast majority of peptide prodrugs tested as novel cancer therapeutics have targeted proteases preferentially or exclusively overexpressed in cancer cells. For instance, the prostate-specific antigen (PSA) and prostate-specific membrane antigen (PSMA) have been popular targets in prostate cancer therapeutics as their expression is relatively restricted and markedly increased in patients with prostate cancer (106-109). Although there have been success stories in the peptide prodrug field, many prodrugs have failed preclinical tests because of premature activation at sites other than the tumor. Redundancy in substrate specificity and non-canonical substrates can lead to non-specific prodrug activation, increasing possible side effects and decreasing efficacy (110). Recent studies have attempted to identify novel human protease and substrate pairs to improve the specificity issues seen with the more common proteases (111). Alternatively, a genetically engineered virus can be used to introduce an exogenous activating enzyme into the cell of choice. This type of therapy is known as virus-directed enzyme prodrug therapy (VDEPT) (112, 113). While not
previously studied, we hypothesized that it would be possible to use a naturally occurring
virally encoded protease, such as SVV-001 \(3\text{C}^{\text{pro}}\), for the same purpose.
**Study Introduction**

One strategy to harness the cancer selective activity of SVV-001 and improve its efficacy is to combine virotherapy with a peptide prodrug selectively activated at the site of infection by a virus-encoded protease, 3C\textsuperscript{pro}. The goal of this study was to identify optimized peptide substrates for SVV-001 3C\textsuperscript{pro} and to demonstrate selective peptide hydrolysis *in vitro* as the basis for developing a novel VDEPT strategy (Figure 1). Instead of using a virus simply as a transport or targeting vector for the enzyme gene, this approach harnesses the intrinsic activity of SVV-001 3C\textsuperscript{pro} already existing in the wild-type SVV-001 proteome. The selectivity of the oncolytic virus for the tumor site and the existence of the active 3C\textsuperscript{pro} produced by the virus obviates the need for genetic engineering of the virus to carry an enzyme gene into the cells.
Figure 1. Conceptual schematic of use of an SVV 3C\textsuperscript{pro} activated peptide prodrug in combination with SVV virotherapy as a novel form of VDEPT. SVV infects a fraction of tumor cells (1), producing 3C\textsuperscript{pro} during the viral life cycle. Upon cell lysis, new SVV virions and 3C\textsuperscript{pro} are released into nearby tissue (2). Administered peptide prodrug would be excluded from cells by the presence of the attached peptide (3), sparing normal tissues, which are non-permissive and therefore cannot express 3C\textsuperscript{pro}. The 3C\textsuperscript{pro} present at high concentration exclusively within the tumor microenvironment cleaves this peptide sequence (4), allowing the cytotoxic moiety to enter both infected and adjacent uninfected cells within the tumor, resulting in a powerful local bystander effect (5).
In this study, we quantitatively characterized the proteolytic cleavage of predicted endogenous SVV-001 3C\textsuperscript{pro} substrates \textit{in vitro} using a FRET fusion protein approach. We discovered essential positions and optimal amino acid residues within one of the endogenous substrate sequences of the SVV polyprotein (61, 114). Once we identified a substrate with high turnover efficiency, we confirmed the ability of the substrate, L/VP4.1, to be recognized and cleaved in a cellular assay by the 3C\textsuperscript{pro} produced during an active SVV-001 infection in the context of both a FRET fusion protein and a fluorogenic peptide. Finally, we determined the stability of the L/VP4.1 substrate in human serum for use in the development of a peptide prodrug. In summary, these studies have identified an optimized substrate to be used as the peptide in the 3C\textsuperscript{pro} activated peptide-prodrug as a novel VDEPT approach.
**Materials and Methods**

**Reagents and Bacterial Strains**

Polymerase chain reactions (PCRs) were carried out using a GeneAmp PCR System 9700 Thermocycler (Invitrogen). PCR fragments were purified using QIAquick PCR Purification Kit (Qiagen). All restriction and ligation enzymes were purchased from New England Biolabs, Inc. Competent DH10B and BL21 AI cells were purchased from Invitrogen. Plasmids were isolated and purified from bacteria using QIAquick Spin Miniprep Kit (Qiagen). Recombinant HRV 3C Protease was purchased from Sigma. Nucleotide and protein sequence alignments were performed in Geneious Pro 4.7.6. The GenBank/EMBL/DDBJ accession number for the complete genome sequence of SVV-001 is DQ641257. Genome sequences of related picornaviruses were obtained from NCBI GenBank (RefSeq. IDs NC_001366.1, NC_001479.1, NC_009448.2, NC_010810.1, NC_011349.1).

**Plasmid Construction**

*FRET Substrates*—The fluorescent FRET protein pair CyPet and YPet were used to construct SVV-001 3C\textsuperscript{pro} substrates. The plasmid pBad33CGSYK, a gift from Dr. Patrick Daugherty, which expresses CyPet and YPet with a C-terminal 6×His tag separated by a flexible linker (SGGSGST), a non-hydrolyzable linker (NHL; GGSGGS), and a second flexible linker (GGGSGGS), was used as a template for constructing SVV-001 3C\textsuperscript{pro} substrates. Using forward primers 1-21 and reverse primer 22 (Table 1) the PCR fragments were amplified to encode the corresponding substrates. Purified PCR fragments were then digested with *KpnI*
and SphI and ligated into a similarly digested pBad33CGSYK to yield circularized plasmids containing each of the substrates flanked by a pair of flexible linkers and the CyPet and YPet proteins. Ligated plasmids were then transformed into competent DH10B cells and clones were selected on LB agar (Sigma) plates supplemented with 34 µg/ml chloramphenicol (Sigma). Substrates were verified by Sanger sequencing.

*SVV-001 3C<sup>pro</sup> Plasmid*— SVV-001 3C<sup>pro</sup> was cloned by fusion to Maltose Binding Protein (MBP) via a Tobacco Etch Virus (TEV) protease sequence (ENLYFQG) in a bacterial expression vector using Gateway Cloning (Invitrogen) to create a 6×His-MBP-TEV-SVV-001 3C<sup>pro</sup> fusion protein plasmid (115-117). The plasmid pNTX-09, expressing the full-length viral cDNA, was used as a template for PCR (62). The PCR fragment was amplified using forward primer 23 and reverse primer 24, incorporating attB1 and attB2 sites for Gateway cloning. The purified PCR fragment was then used in a BP reaction (Invitrogen) with pDONR221, a destination Gateway plasmid with a kanamycin resistance gene. The recombinant plasmid was then transformed into DH10B cells and clones were selected on LB agar plates supplemented with 50 µg/ml kanamycin (Sigma). The purified plasmid containing the PCR fragment was used in a LR reaction (Invitrogen) with pDEST566 (Addgene plasmid #11517) (117). The plasmid expresses an N-terminal 6×His MBP. The recombinant plasmid was then transformed into DH10B cells and clones were selected on LB agar plates supplemented with 100 µg/ml carbenicillin (Sigma). The purified pDEST566 plasmid containing the PCR fragment (6×His-MBP-TEV-SVV-001 3C<sup>pro</sup>) was then transformed into BL21 AI cells and clones were selected on LB agar plates supplemented with 100 µg/ml carbenicillin. The catalytic dead mutant 3C<sup>pro</sup>, SVV-001 C160A 3C<sup>pro</sup>, was cloned using QuickChange site-directed mutagenesis (Stratagene) using primers 25 and 26.
and 6×His-MBP-TEV-SVV-001 3C<sup>pro</sup> plasmid described above. The template DNA was digested by DpnI and the mutant plasmid was transformed into DH10B cells and subsequently into BL21 cells as described above. The protease sequence and the C160A mutation were confirmed using Sanger sequencing.
Table 1. Oligonucleotide list for primers used for the construction of FRET fusion proteins and recombinant SVV-001 3C<sup>pro</sup>.

<table>
<thead>
<tr>
<th>Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-HRV</td>
<td>5'-GGGGTACCTCGGAGTTCTGTTCCAGGGTGCTCCGGGTGGTAGCTGCAATCGGGGTGG-3’</td>
</tr>
<tr>
<td>2-L/VP4</td>
<td>5'-GGGGTACCACTCGTTCAGGAACAGCCGGTCCGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>3-VP4/VP2</td>
<td>5'-GGGGGTACCGATCGTTATGGTTACGAACTGCAGGGTCCGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>4-VP2/VP3</td>
<td>5'-GGGGGTACCGATCGTTATGGTTACGAACTGCAGGGTCCGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>5-VP3/VP1</td>
<td>5'-GGGGTACCTCGGAGTTCTGTTCCAGGGTGTAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>6-VP1/2A</td>
<td>5'-GGGGTACCTCGGAGTTCTGTTCCAGGGTGTAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>7-2B/2C</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>8-2C/3A</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>9-3A/3B</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>10-3B/3C</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>11-3C/3D</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>12-L/VP4.1</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>13-2B/2C.1</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>14-L/VP4.2</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>15-L/VP4.3</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>16-L/VP4.4</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>17-L/VP4.5</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>18-L/VP4.6</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>19-L/VP4.7</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>20-L/VP4.8</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>21-L/VP4.9</td>
<td>5'-GGGGTACCCACCCCTTCTGCAGCAATCGGGGTGGTAGCTGAGCAGGGGTGG-3’</td>
</tr>
<tr>
<td>22-pB33CGSYR</td>
<td>5'-ACATGCATGCGGCCACCTT-3’</td>
</tr>
<tr>
<td>23-3CProtease F</td>
<td>5'-GGGGTACCTCGGAGTTCTGTTCCAGGGTGCTCCGGGTGGTAGCTGCAATCGGGGTGG-3’</td>
</tr>
<tr>
<td>24-3CProtease R</td>
<td>5'-GGGGTACCTCGGAGTTCTGTTCCAGGGTGCTCCGGGTGGTAGCTGCAATCGGGGTGG-3’</td>
</tr>
<tr>
<td>25-C160A3CProtease F</td>
<td>5'-TACAAAGGATGGGCGGCTCGGCCCCCCCTG-3’</td>
</tr>
<tr>
<td>26-C160A3CProtease R</td>
<td>5'-GATGTTTCTACCCCGGCAGCAAGGGGCTGCGGTGGTAGCTGCAATCGGGGTGG-3’</td>
</tr>
</tbody>
</table>

Numbers correspond to the primer identity in the main text. Primer name describes the product the primer will eventually create. Oligonucleotides shown from 5’ to 3’.
Protein Induction and Purification

**FRET Substrates**- Overnight cultures of DH10B cells transformed with FRET substrate expression vectors were diluted 1:50 and grown at 37 °C for 3 hr in LB medium (Sigma) supplemented with 34 µg/ml chloramphenicol. Expression of FRET proteins was induced with 0.1% wt/vol L(+)-arabinose (Sigma) for 16 hr at room temperature. Cells were harvested by centrifugation and soluble protein was then isolated using B-PER Protein Extraction Reagent (Pierce). Fusion proteins, which contain a C-terminal 6×His tag, were then purified using HisPur Ni-NTA resin (Pierce) and eluted in Phosphate Buffered Saline (PBS, Quality Biological), pH 7.4 with 150 mM imidazole (Sigma).

**SVV-001 3Cpro Protein**- Overnight cultures of BL21AI cells transformed with pDEST566 expression vector containing the 6×His-MBP-TEV-SVV-001 3Cpro and 6×His-MBP-TEV-SVV-001 C160A 3Cpro fusion protein were diluted 1:100 and grown at 37 °C until the culture reached mid-log growth phase in LB media supplemented with 100 µg/ml carbenicillin. Protein expression was induced via the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG, Thermo Fisher Scientific) and 0.2% wt/vol L(+)-arabinose for 8 hr at room temperature. Cells were harvested and soluble proteins were isolated as above. The fusion proteins were then purified using HisPur cobalt resin (Pierce) and eluted in PBS, pH 7.4 with 150 mM imidazole. The fractions containing the desired proteins were pooled into a 10K MWCO Slide-A-Lyzer dialysis cassette (Pierce) and dialyzed overnight at 4 °C against 25 mM Tris-HCl (Thermo Fisher Scientific) pH 7.5, containing 50 mM NaCl (Thermo Fisher Scientific) and 1 mM dithiothreitol (DTT, Thermo Fisher Scientific). Dialyzed proteins were then concentrated using Amicon Ultra-15 50K MWCO centrifugal
filter units (Millipore). Protein concentrations were determined using the Micro BCA Protein Assay (Pierce) and aliquots were frozen at -80 °C after addition of glycerol to 10% (Sigma). Overexpression and purifications were monitored by SDS PAGE (Figure 2). The purity of the $6\times$His-MBP-TEV-SVV-001 $3\text{C}^{\text{pro}}$ preparation was estimated to be 91% using a GS-800 Calibrated Densitometer (BioRad).
Figure 2. SDS-PAGE gel of overexpression and purification of recombinant fusion SVV-001 3C\textsuperscript{pro} and catalytically dead mutant SVV-001 C160A 3C\textsuperscript{pro}. Lane 1- Molecular weight protein marker. Lane 2- Insoluble protein after cell lysis (1:10 dilution). Lane 3- Soluble protein after cell lysis. Lane 4- Pool eluted fractions from HisPur\textsuperscript{™} cobalt resin purification of 3C protease. Lane 5- Amicon Ultra-15 50K MWCO flow through. Lane 6- Concentrated elution fractions after Amicon Ultra-15. The band corresponding to the recombinant fusion SVV-001 3C\textsuperscript{pro} and SVV-001 C160A 3C\textsuperscript{pro} is labeled.
**In Vitro Cleavage Assays**

*FRET Substrate Cleavage Assay and Kinetic Data Analysis* - Reactions were performed in PBS, pH 7.4 and 200 nM FRET substrate in a total volume of 100 µL. Substrates were incubated with 250 nM SVV-001 3C<sup>pro</sup> at 30 °C for 3 hr. Human rhinovirus (HRV) control substrate was incubated with recombinant HRV 3C<sup>pro</sup> under the same conditions. Fluorescence emissions at 475 nm and 527 nm, corresponding to CyPet and YPet respectively, were followed after excitation at 433 nm every 180 s by a Safire fluorimeter (Tecan). All experiments were performed in triplicate. FRET ratios (YPet fluorescence/CyPet fluorescence) were calculated at each timepoint. Conversion of each substrate to cleaved substrate was calculated by dividing the change in the FRET ratio at each timepoint by the total change in the FRET ratio corresponding to complete cleavage of the substrate. Using the equation from Boulware et al., conversion data was plotted vs. time (s) and fit with the equation:

\[
\text{Conversion} = 1 - \exp\left(-\frac{k_{\text{cat}}}{K_M}[E] \cdot t\right)
\]

where [E] is the 3C<sup>pro</sup> concentration (M) and \(t\) is time (s) (118). The second order rate constant (\(k_{\text{cat}}/K_M\)) was determined by the curve fit using GraphPad Prism software. Reported values are the average \(k_{\text{cat}}/K_M\) values of 5-7 experiments. Uncertainty is expressed by standard deviation. Error bars on data points were removed for figure clarity.

*Fluorescently Quenched Peptide Cleavage Assay and Kinetic Data Analysis* - Peptides corresponding to the substrate sequences L/VP4.1 and non-hydrolysable linker (NHL) were synthesized between a 5-carboxyfluorescein (5-FAM) fluorophore/CPC Quencher (CPQ2)
quencher pair (CPC Scientific). Additional lysines were added for increased solubility giving final peptide sequences of CPQ2-IVYELQGP-K(5FAM)-KK-NH2 and CPQ2-GGSGGS-K(5FAM)-KK-NH2 for L/VP4.1 and NHL, respectively. The identity and sequence of the L/VP4.1 and NHL peptides were confirmed by CPC Scientific via LC-MS and determined to be 96.7% and 95.1% pure, respectively. The peptides were diluted in H2O to concentrations ranging from 7 µM to 300 nM and incubated with 50 nM purified SVV-001 3Cpro for 1 hr at 37 °C. Fluorescence emissions at 520 nm, corresponding to the liberated 5-FAM fluorophore peptide cleavage product was followed after excitation at 490 nm by a DTX 880 multimode detector (Beckman Dickinson) every 60 s. Experiments were performed in triplicate in a 96 well flat bottom black opaque plate. Standard curves of 5-FAM fluorescence vs. concentration were generated to convert the relative fluorescence units to moles of cleaved product generated during the reaction. Initial rates of hydrolysis were calculated from data during the first 3 minutes of the reaction. The initial rates of reaction from 3-4 separate experiments were then averaged and used to calculate kinetic rate constants. Standard deviation values were calculated from 3-4 separate experiments. Kinetic constants $K_M$ and $V_{max}$ were calculated by plotting initial rates of hydrolysis vs. substrate concentration and fitting plots with the Michaelis-Menten equation using GraphPad Prism software. Using the equation $V_{max} = k_{cat} \times [E_t]$ and solving for $k_{cat}$, we were then able to determine the second order rate constant, $k_{cat}/K_M$.

**Cellular Assays**

**SVV-001 Cellular Assay**- All SVV-001 stocks were cultured and purified as described previously and virus titers were determined by tissue culture infective dose (TCID$_{50}$) (119).
The SCLC cell line, NCI-H446 (ATCC) was cultured in RPMI 1640 media (Quality Biological) supplemented with 10% fetal bovine serum (HyClone). The cell line NCI-H446 is routinely tested and authenticated by short tandem repeat (STR) analysis by DDC Medical and was last authenticated six months before submission.

**FRET Substrate Cellular Assay**—Cells ($1.6 \times 10^7$) were plated on a 150 mm cell culture dish (Corning) and allowed to recover overnight at 37 °C. Each plate was inoculated with SVV-001 at the TCID$_{50}$ and infected cells were incubated at 37 °C for 8 hr. After aspiration of the media, cells were mechanically lifted from the plate in phenol-red free RPMI 1640 media (Gibco) and pelleted by centrifugation. The cell pellet was then resuspended in phenol-red free RPMI 1640 media and aliquoted into the wells of a 96-well flat bottom black opaque plate (Corning). FRET substrates, L/VP4.1 or NHL, were added at a final concentration of 200 nM and incubated for 3 hr at 37 °C. During the substrate incubation the change in FRET was measured with a Tecan Safire fluorimeter as described above. Conversion of the L/VP4.1 substrate to cleaved substrate was calculated by dividing the change in the FRET ratio at each time point by the total change in the FRET ratio corresponding to complete cleavage of the substrate. The conversion of the NHL substrate was similarly determined by dividing the change in the FRET ratio of the NHL at each time point by the total change in the FRET ratio of the L/VP4.1 substrate corresponding to complete cleavage. Using the previously calculated second order rate constant for L/VP4.1 and the conversion equation described previously, the concentration of SVV-001 3C$^\text{pro}$ was determined ($118$). The reported values are averages of 7 experiments. Uncertainty is expressed by standard deviation. Error bars on data points were removed for figure clarity.
**Fluorescently Quenched Peptide Cellular Assay**- A similar cellular assay was performed as described in the previous paragraph with the exception of using the CPQ2/5-FAM L/VP4.1 and NHL fluorogenic peptides described earlier. After an 8 hour infection with SVV-001, NCI-H446 cells were removed from the cell culture dish using phenol-red free RPMI 1640 media supplemented with 0.1% ethylenediaminetetraacetic acid (EDTA, Sigma) to aid in gentle cell detachment, and pelleted via centrifugation. Pellets were resuspended in phenol-red free RPMI 1640 prior to plating in black opaque 96 well plates as above. CPQ2/5-FAM peptides, L/VP4.1 or NHL were added at a final concentration of 3 µM to plated cells and incubated for 3 hours at 37 °C. Fluorescence emissions at 520 nm, corresponding to the liberated 5-FAM fluorophore peptide cleavage product was followed after excitation at 490 nm by a Synergy Neo multimode plate reader (Biotek) every 60 s. The experiment was performed in triplicate in a 96 well flat bottom black opaque plate and average fluorescence units were used to observe changes in fluorescence. Standard curves of 5-FAM fluorescence vs. concentration were generated to convert the relative fluorescence units to moles of cleaved product generated during the reaction. Initial rates of hydrolysis were calculated from data during the first 10 minutes of the reaction and averaged from 3-4 separate experiments. Uncertainty is expressed by standard deviation and shown with error bars.

**Plasma Stability Assay**

Human plasma was obtained from pooled discarded clinical samples less than 6 hrs old. Human plasma was isolated from whole blood by centrifugation, pooled, and diluted 1:1 with distilled H₂O (Gibco). The L/VP4.1 and NHL fluorogenic peptides (20 µM) were incubated with plasma for 1 hr at 37 °C. Fluorescence emissions were obtained as described above by a
DTX 880 multimode detector (Beckman Dickinson) every 60 s. Experiments were performed in triplicate in a 96 well flat bottom black opaque plate and the average fluorescence units were used to observe changes in fluorescence. Uncertainty is expressed by standard deviation.
Results

Based on sequence alignment to closely related cardioviruses, the SVV-001 polyprotein has been predicted to consist of twelve proteins after cleavage (Figure 3A). While cleavage sites for the structural proteins VP1-4 have been confirmed by N-terminal sequencing, other putative 3C\textsuperscript{pro} cleavage sites have been proposed based entirely on sequence homology (Figure 3B) (61). The consensus Gln-Gly-Pro cleavage motif (Q↓GP) found for other picornaviruses is present in many predicted substrates; however, cleavage sites without this consensus sequence are also predicted (61). As the endogenous sequences were expected to have the highest probability of being efficiently cleaved by the SVV-001 3C\textsuperscript{pro}, our initial kinetic studies focused on the ten intrinsic substrates to identify the substrate with the highest turnover efficiency.

To identify an optimized peptide sequence for cleavage by SVV-001 3C\textsuperscript{pro}, two approaches to measure peptide cleavage were pursued. The first approach utilized FRET as a reporter for protease-catalyzed peptide cleavage. Each of the ten predicted viral cleavage sites were cloned between the FRET donor/acceptor pair, CyPet and YPet, to create FRET fusion protein substrates (114, 118, 120). In this context, FRET is dependent on maintaining close proximity of the CyPet and YPet pairs: cleavage of the fusion proteins by purified SVV-001 3C\textsuperscript{pro} should result in loss of FRET (Figure 3C). The decrease in FRET was measured in real time using a fluorimeter, and second order rate constants ($k_{cat}/K_M$) were determined as described previously (118). We chose an 8 residue window spanning P6-P2’ for all studies based on structural and functional studies of other picornavirus 3C proteases implicating a greater importance of non-prime side residues on substrate turnover efficiency (90, 93).
SVV-001 3C\textsuperscript{pro} was cloned and purified as a recombinant fusion construct with a 6x Histagged maltose binding protein (MBP). A non-hydrolysable linker FP substrate (NHL; GGSGGS) and an optimized human rhinovirus 3C\textsuperscript{pro} FP substrate (HRV; LEVLFQGP) were used as negative and positive controls, respectively, the latter being incubated with recombinant HRV 3C\textsuperscript{pro} (93, 118). Interestingly, only two of the ten endogenous substrates, L/VP4 and 2B/2C, were efficiently cleaved by SVV-001 3C\textsuperscript{pro} \textit{in vitro} in the context of a FP substrate. All other substrates failed to show measurable cleavage over 3 hours of incubation with the protease at 30 °C. The kinetics of substrate hydrolysis or conversion were calculated (Figure 3D). Both the 2B/2C and L/VP4 FP substrates were determined to have similar $k_{cat}/K_M$ values at $1204 \pm 146$ M\textsuperscript{-1}s\textsuperscript{-1} and $1932 \pm 183$ M\textsuperscript{-1}s\textsuperscript{-1}, respectively. The optimized HRV FP substrate displayed the highest turnover efficiency ($k_{cat}/K_M = 5079 \pm 528$ M\textsuperscript{-1}s\textsuperscript{-1}), which was ~3-4-fold higher than either SVV FP substrate (Table 2). Incubations of the 2C/3A, 3B/3C, and 3C/3D FP substrates with an increased protease concentration of 1 µM over three hours did demonstrate slow but nearly complete cleavage (Figure 4). All other FP substrates showed no measurable cleavage by higher concentrations of SVV-001 3C\textsuperscript{pro}.
Figure 3. SVV-001 polyprotein and identification of efficient 3C protease substrates

A. The SVV-001 polyprotein is hypothesized to include twelve mature proteins based on sequence alignment with closely related viruses. Protein regions (drawn to scale) are presented with proposed 3C<sup>pro</sup> cleavage sites (arrowheads) depicted. The 2A ribosome skipping sequence (RS; diamondhead arrow) is also shown.

B. Sequence alignment of proposed 3C<sup>pro</sup> cleavage sites. Sites are named based on the two mature proteins flanking these sites. The presumed scissile bond is depicted with an arrowhead with the consensus Q↓GP cleavage sequence amino acids highlighted.

C. Schematic of FRET substrate construction and kinetic assays. Substrates are cloned between two fluorescent proteins, CyPET and YPET. These molecules exhibit FRET when in close proximity; therefore, there will be a high level of emission at 527 nm (YPET) and lower emission at 475 nm (CyPET). As the 3C protease cleaves and releases YPET from proximity to CyPET, the amount of FRET will decrease observed as an increase in CyPET emission (475 nm) and a decrease in YPET emission (527 nm).

D. Conversion of FRET substrates by purified HRV or SVV-001 3C<sup>pro</sup> over time. Data points represent the average of three replicates at each time point. The HRV substrate was used as a positive control. L/VP4 and 2B/2C are endogenous substrates. L/VP4.1 (P2' N→P substitution) is a further optimized version of the endogenous L/VP4 substrate. Lines of the same color correspond to curve fits from GraphPad. Error bars on data points were removed for figure clarity.
Table 2. Summary of second order rate constants ($k_{cat}/K_M$) for the two endogenous substrates, controls and L/VP4 amino acid substitution/truncation mutants.

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Sequence</th>
<th>$k_{cat}/K_M$ ($M^{-1}s^{-1}$)</th>
<th>SD</th>
<th>Relative to WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>L/VP4</td>
<td>IVYELQGN</td>
<td>1932</td>
<td>183</td>
<td>---</td>
</tr>
<tr>
<td>2B/2C</td>
<td>KLFKMQGP</td>
<td>1204</td>
<td>146</td>
<td>---</td>
</tr>
<tr>
<td>HRV</td>
<td>LEVLFQGP</td>
<td>5079</td>
<td>528</td>
<td>---</td>
</tr>
<tr>
<td>NHL</td>
<td>GGSGGS</td>
<td>N/A</td>
<td>N/A</td>
<td>---</td>
</tr>
<tr>
<td>L/VP4.1</td>
<td>IVYELQGP</td>
<td>17446</td>
<td>2203</td>
<td>9.0</td>
</tr>
<tr>
<td>L/VP4.2</td>
<td>IVYE PQGP</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>L/VP4.3</td>
<td>IVFELQGP</td>
<td>12160</td>
<td>2157</td>
<td>6.3</td>
</tr>
<tr>
<td>L/VP4.4</td>
<td>IVME PQGP</td>
<td>10344</td>
<td>1186</td>
<td>5.3</td>
</tr>
<tr>
<td>L/VP4.5</td>
<td>VYELQGP</td>
<td>9839</td>
<td>696</td>
<td>5.1</td>
</tr>
<tr>
<td>L/VP4.6</td>
<td>YELQGP</td>
<td>1939</td>
<td>304</td>
<td>1.0</td>
</tr>
<tr>
<td>L/VP4.7</td>
<td>ELQGP</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>L/VP4.8</td>
<td>LQGP</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

The altered substrates are shown with the amino acid(s) substitution(s) or truncations shown in bold. Reported values are the average $k_{cat}/K_M$ values of 5-7 experiments. Uncertainty is expressed by standard deviation.
Figure 4. Conversion of Substrates after Incubation with Increased Concentration of SVV-001 3Cpro. All substrates except the L/VP4 and 2B/2C were incubated with 1 µM purified SVV-001 3Cpro for 3 hours at 30 °C. The decrease in FRET was measured in real time using a fluorimeter and data was converted to display fraction of substrate converted to hydrolyzed substrate over time. Data points represent the average of three replicates at each time point. The HRV 3Cpro substrate was used as a positive control. 2C/3A, 3B/3C, and 3C/3D are endogenous substrates that showed cleavage. All other substrates were not cleaved in the presence of increased SVV-001 3Cpro (data not shown).
Surprisingly, of the two efficiently cleaved substrates, only the 2B/2C substrate contains the consensus Q↓GP motif typically found in picornavirus 3C<sup>pro</sup> substrates. The L/VP4 substrate has a variant Q↓GN cleavage site, which did not appear to cause deleterious effects on the kinetics of hydrolysis. To investigate the contribution of the P2’ amino acid position, we exchanged the P2’ amino acid of 2B/2C and L/VP4 to create 2B/2C.1 (2B/2C with Q↓GN) and L/VP4.1 (L/VP4 with Q↓GP, Table 2). These new synthetic FP substrates were incubated with SVV-001 3C<sup>pro</sup> and the kinetics of conversion were calculated as described above (Figure 3C and Table 2). While substituting Asn for Pro in the P2’ position of the 2B/2C substrate abolished hydrolysis, the P2’ position Pro substitution in the L/VP4 substrate increased the $k_{cat}/K_M$ to $17446 \pm 2203$ M<sup>-1</sup>s<sup>-1</sup>, an approximately 9-fold increase. We mutated the catalytic cysteine residue of SVV 3C<sup>pro</sup> to an alanine (C160A 3C<sup>pro</sup>) to eliminate any proteolytic activity and confirmed the change in FRET of the FP substrates was entirely due to the activity of the 3C<sup>pro</sup> (Figure 5).
Figure 5. Conversion of Substrates after Incubation with SVV-001 3C<sup>pro</sup> compared to SVV-001 C160A 3C<sup>pro</sup>. Substrates L/VP4 and L/VP4.1 were incubated with 250 nM purified SVV-001 3C<sup>pro</sup> (labeled WT) or SVV-001 C160A 3C<sup>pro</sup> (labeled C160A) for 3 hours at 30 °C. The decrease in FRET was measured in real time using a fluorimeter and data was converted to display fraction of substrate converted to hydrolyzed substrate over time. Changes in FRET from substrates incubated with C160A 3C<sup>pro</sup> were divided by total change in FRET by substrates incubated with WT 3C<sup>pro</sup> indicating complete hydrolysis. Data points represent the average of three replicates at each time point.
To assess the specificity of L/VP4.1, we tested the capacity for the proteases from SVV, HRV, and Coxsackie A virus (CAV) to cleave one another’s native substrates. Interestingly, while cross-selectivity was observed between HRV and CAV, we observed no turnover of L/VP4.1 by any non-native 3C protease (Figure 6). Additionally, we tested the ability of the SVV 3Cpro to cleave the alternative cleavage sequence motif Q↓SP by cloning a P1’ GN→SP substituted form of L/VP4 and confirmed it had an increased turnover efficiency over L/VP4, but was inferior to L/VP4.1 (L/VP4.9; Figure 7).
Figure 6. Comparison of SVV 3C\textsuperscript{pro}-L/VP4.1 Substrate Specificity to Selected Picornavirus Protease-Substrate Pairs. A. Substrates L/VP4.1 (labeled SVV), HRV, and Coxsackievirus (CAV) 2C/3A substrate (labeled CAV) with amino acid sequence MEALFQ↓GP were incubated with 250 nM SVV 3C\textsuperscript{pro} (black data points), HRV 3C\textsuperscript{pro} (blue data points), or CAV 3C\textsuperscript{pro} (red data points) for 3 hours at 30 °C. CAV 3C\textsuperscript{pro} was cloned, overexpressed, and purified using the same methods described for SVV 3C\textsuperscript{pro} in the Methods section. The decrease in FRET was measured in real time using a fluorimeter and data was converted to display fraction of substrate converted to hydrolyzed substrate over time. Data points represent the average of three replicates at each time point. B. Heat map depicting relative rates of cleavage for protease-substrate pairs compared to the native protease-substrate pair.
Figure 7. Comparison of Conversion of Substrate containing Q↓SP cleavage site to L/VP4 and L/VP4.1 Substrates. Substrate L/VP4.9 with amino acid sequence IVYELQ↓SP was cloned between the FRET pair, CyPET and YPET using primer numbers 21-22 (Table S1), induced for overexpression, and purified using methods described for FRET substrates in the main text. Substrates L/VP4, L/VP4.1, and L/VP4.9 were incubated with 250 nM purified SVV-001 3Cpro for 3 hours at 30 °C. The decrease in FRET was measured in real time using a fluorimeter and data was converted to display fraction of substrate converted to hydrolyzed substrate over time. Data points represent the average of three replicates at each time point.
Although the P2’ Pro substitution (L/VP4.1) significantly increases $k_{cat}/K_M$, the P2’ Asn is universally conserved in the VP4 proteins of cardioviruses. Amino acids in the non-prime P1-6 sites of the L/VP4 sequence are more divergent (Figure 8A). Thus, the non-prime L/VP4.1 amino acids may be evolutionarily optimized for rapid cleavage under the constraint of a suboptimal P2’ Asn. To determine substrate requirements in these positions, we substituted amino acids found in related cardioviruses in the P2 and P4 sites to create L/VP4.2, (P2 L→P substitution), L/VP.3 (P4 Y→F substitution), and L/VP4.4 (P4 Y→M substitution) (Figure 8B). We also used the L/VP4.1 substrate to determine substrate length constraints for the protease by sequentially removing P6-P3 amino acids from the non-prime side and testing all truncated substrates for cleavage (L/VP4.5-.8; Figure 8C).
Figure 8. Optimization of the L/VP4 Substrate  

A. Alignment of SVV L/VP4 cleavage site (shown in italics) with L/VP4 junction sequences of closely related cardioviruses. The amino acids in each sequence that diverge from the consensus sequence (shown in bold) are highlighted.  

B. Conversion of modified L/VP4.1 FRET substrates by purified SVV-001 3Cpro over time. Data points represent the average of three replicates at each time point. The endogenous L/VP4 substrate and optimized L/VP4.1 substrate are shown for reference. L/VP4.3 and L/VP4.4 are P4 Y→F substitution and P4 Y→M substitution substrates, respectively. Lines of the same color correspond to curve fits from GraphPad. Error bars on data points were removed for figure clarity.  

C. Conversion of truncated L/VP4.1 FRET substrates by purified SVV-001 3Cpro over time. Data points represent the average of three replicates at each time point. The endogenous L/VP4 substrate and optimized L/VP4.1 substrate are shown for comparison. L/VP4.5 and L/VP4.6 are P6 and P5/P6 truncations of L/VP4.1, respectively. Lines of the same color correspond to curve fits from GraphPad. Error bars on data points were removed for figure clarity.
The FRET ratios from the experiment were used to measure conversion and calculate \( k_{cat}/K_M \) (Table 2). The substitution made to create L/VP4.2 was detrimental, as this construct was not efficiently cleaved in 3 hours. The L/VP4.3 and L/VP4.4 FP substrates did undergo cleavage more efficiently than the endogenous L/VP4 FP substrate, but 30% and 40% less efficiently than the L/VP4.1 Pro variant, respectively. Removal of the P6 Ile residue in L/VP4.5 decreased the turnover efficiency of cleavage from L/VP4.1 by almost 2-fold but remained a better substrate than the endogenous L/VP4. Removal of both P6 and P5 amino acids (L/VP4.6) decreased the \( k_{cat}/K_M \) to values similar to the endogenous L/VP4 sequence. Further truncation of the substrate abolished detectable cleavage by SVV-001 3C\textsuperscript{pro} (Table 2).

Determination of an optimized amino acid sequence and length for efficient recognition and proteolysis described above was performed using the recombinant SVV-001 3C\textsuperscript{pro}. To verify that the lead substrate would be actively cleaved by the native 3C\textsuperscript{pro}, the L/VP4.1 FP substrate was evaluated as a substrate for proteolysis catalyzed by SVV-001 3C\textsuperscript{pro} produced during an active SVV-001 infection, in which 3C\textsuperscript{pro} released during a lytic viral infection is the sole source of the 3C\textsuperscript{pro}. Substrates were incubated for 3 hours at 37 °C with the SCLC cell line, NCI-H446, after an 8 hour infection with SVV-001 (Figure 9A). Eight hours after infection, cells infected with SVV undergo cell lysis releasing both new SVV virions as well as 3C\textsuperscript{pro} (62). Loss of FRET, indicative of cleavage, was only observed when the L/VP4.1 FP substrate was incubated with SVV-001 infected H446 cells. Incubations of either FP substrate with uninfected cells, or infected cells with the NHL FP substrate, did not show a loss of FRET over 3 hours. With the second order rate constant previously calculated for the L/VP4.1 FP substrate, we determined the concentration of 3C\textsuperscript{pro} during the SVV infection to be 30.4 ± 4.9 nM using the equation described in Boulware et al (118). Since loss of FRET
was observed exclusively in the presence of infected cells, we conclude that proteolysis of the L/VP4.1 FP substrate is catalyzed selectively by SVV-001 3C\textsuperscript{pro} and not cellular proteases expressed by H446 cells.
Figure 9. SVV-001 3C<sup>pro</sup> substrates are cleaved in the context of a cellular infection. A. Conversion of FRET substrates by SVV-001 3C<sup>pro</sup> produced by a cellular SVV infection of permissive SCLC line, NCI-H446. Data points represent the average of three replicates at each time point. The L/VP4.1 FP substrate incubated with uninfected and infected cells, using similar incubations of NHL FP substrate as a negative control. Lines of the same color correspond to curve fits from GraphPad. Error bars on data points were removed for figure clarity. B. Initial reaction rates of L/VP4.1 peptide cleavage by recombinant SVV-001 3C<sup>pro</sup>. Data points represent the initial rate of reaction at each concentration of L/VP4.1 peptide calculated from three replicate experiments. Data points were fit to a Michaelis-Menten nonlinear regression from GraphPad and the kinetic constants determined by the curve fit were reported. Standard deviation values of the kinetics constants were calculated by the GraphPad software and propagated through second order rate constant calculations. C.
Proteolysis of CPQ2/5-FAM peptides by native SVV-001 3C<sup>pro</sup> in a cellular assay with NCI-H446. Data points represent the average relative fluorescence units (RFUs) increase of three replicates at each time point relative to fluorescence at time zero. The L/VP4.1 FQ peptide was incubated with uninfected and infected cells, using similar incubations of NHL FQ peptide as a negative control. Lines of the same color correspond to connecting line between points.
We considered the possibility that the FRET fusion peptide substrate for SVV-001 3C\textsuperscript{pro} could influence peptide cleavage kinetics. Thus, a second approach to determine the efficiency of cleavage of L/VP4.1 by SVV-001 3C\textsuperscript{pro} employed a 5-FAM/CPQ2 fluorophore/quencher pair incorporated into the peptide (Figure 9B). In this case, SVV-001 3C\textsuperscript{pro} catalyzed peptide cleavage will result in an increase in fluorescence as 5-FAM is released through proteolysis. The peptide substrate, synthesized between a 5-FAM/CPQ2 fluorophore/quencher pair, was incubated at concentrations ranging from 7 µM to 500 nM with 50 nM recombinant 3C protease for 1 hour at 37 °C. The NHL substrate was also synthesized as an FQ peptide and used as a negative control. From the initial rates of reaction, we determined the second order rate constant ($k_{cat}/K_M$) for the L/VP4.1 peptide to be $5.42 \pm 1.17 \times 10^5 \text{ M}^{-1}\text{s}^{-1}$. The difference in specificity constants may be attributable to the different contexts in which the substrate sequence is presented. In the L/VP4.1 FP substrate, two large proteins flank the relatively short substrate, which could be limiting access of the SVV-001 3C\textsuperscript{pro} despite the presence of a flexible linker. Conversely, the fluorophore and quencher in the L/VP4.1 FQ peptide is much smaller in size and therefore may better reflect the context of the L/VP4.1 substrate in the native SVV-001 polyprotein.

The L/VP4.1 FQ peptide was also tested for proteolysis by the native 3C\textsuperscript{pro} produced during a cellular SVV infection. Using the NHL FQ peptide as a negative control, the substrates were incubated at 37 °C for 3 hours with NCI-H446 cells, following an 8 hour infection with SVV-001 (Figure 9C). Similar to the cellular assay with FP substrates, cells infected with SVV are undergoing cell lysis 8 hrs post infection, releasing both new SVV virions as well as 3C\textsuperscript{pro} (62). A significant increase in fluorescence, indicative of cleavage of the FQ peptide, was only observed in the incubation of the L/VP4.1 FQ peptide with SVV-001 infected cells.
The NHL FQ peptide was not cleaved in incubations with either uninfected or infected H446 cells. The L/VP4.1 FQ peptide incubated with uninfected cells also did not show an appreciable increase in fluorescence. The initial rates of reaction were determined to be 0.168 ± 0.0025 pmol/sec and 0.446 ± 0.0056 pmol/sec for the L/VP4.1 FQ peptide incubated with uninfected and SVV infected cells, respectively. As with the cellular FRET assay, we conclude that the L/VP4.1 FQ peptide is selectively cleaved by SVV-001 3C\textsuperscript{pro}.

A peptide prodrug based on the L/VP4.1 substrate would most likely be administered intravenously, therefore it was important to determine the stability of the peptide in human plasma (Figure 10). Both the L/VP4.1 and NHL FQ peptides were incubated at a final concentration of 20 µM with a 50% plasma/water solution for 1 hour at 37 °C. There was no significant change in fluorescence for either the NHL or L/VP4.1 FQ peptide relative to plasma alone. Therefore, we conclude that the L/VP4.1 FQ peptide is stable in human plasma.
Figure 10. Stability of L/VP4.1 fluorescently quenched peptide in human plasma. Data points represent the average relative fluorescence units (RFUs) increase of four replicates at each time point relative to fluorescence at time zero for L/VP4.1 FQ peptide or NHL FQ peptide incubated with human plasma. The peptides were also incubated with human plasma supplemented with SVV 3C\textsuperscript{pro} or SVV C160A 3C\textsuperscript{pro} as positive and negative controls, respectively. Uncertainty is expressed by standard deviation.
Discussion

Proteases participate in a large number of essential functions, including but not limited to DNA transcription, cell proliferation, signaling cascades, microenvironment remodeling, and cell death (121). Their roles in these processes depend on the intrinsic activity of the protease and its ability to selectively cleave its substrate(s). Because of their innate specificities, many cellular proteases have been used to selectively activate peptide prodrugs. The regulated release of the active cytotoxic agent is designed to reduce toxic side effects of the parent drug while increasing its specificity for the targeted cells (105). Although there have been success stories in the peptide prodrug field, many prodrugs have failed preclinical tests because of premature activation at sites other than the tumor. Redundancy in substrate specificity and non-canonical substrates can lead to non-specific prodrug activation, increasing possible side effects and decreasing efficacy (110).

Anticipating these issues, we focused on designing an optimized substrate that was efficiently and specifically cleaved by SVV-001 3C\textsuperscript{pro}. The use of the FP substrates and FQ peptides allowed us to follow kinetic assays in real-time with a purified recombinant form of the 3C\textsuperscript{pro} as well as the native SVV-001 3C\textsuperscript{pro} produced during a cellular infection. The FRET protein substrates had the added advantage of being easily modified by cloning techniques, allowing us to test and identify multiple optimized substrates that were specifically cleaved by the protease (118, 122). Cellular experiments with the L/VP4.1 FP substrate and FQ peptide ultimately determined the specificity of the SVV-001 3C\textsuperscript{pro} for the L/VP4.1 substrate produced during an SVV-001 infection and allowed us to estimate the total 3C\textsuperscript{pro} concentration in extracellular space.
We have shown that the recombinant fusion SVV-001 3C\textsuperscript{pro} rapidly cleaved two of the ten proposed 3C\textsuperscript{pro} substrates. Efficient substrate recognition and hydrolysis by many proteases is not solely based on the primary amino acid sequence; secondary and in some cases, tertiary structures also play a role in regulating proteolytic cleavage. For a majority of picornaviral polyproteins, protein precursors begin to undergo folding before cleavage between the individual proteins even occurs (96, 123). In the context of an isolated protease substrate, important structural features of the polyprotein are lost, which could substantially alter or even abolish substrate hydrolysis. However, a ribosome skipping event following the 2A peptide positions both the L/VP4 and 2B/2C sites at the amino terminal ends of nascent peptides, making them easily accessible to SVV-001 3C\textsuperscript{pro}, and potentially reducing complex secondary and tertiary structure (61, 102). Their intrinsic lack of rigid structure may explain why only these two substrates were efficiently cleaved by SVV-001 3C\textsuperscript{pro} in the context of the FP substrates.

Viral fusion proteins, intermediates in the proteolytic cascade, may also play a role in efficiency of substrate recognition and cleavage. Fusion proteins such as 3CD may be responsible for some of the cleavage events in the SVV-001 polypeptide (124, 125). We have not ruled out the possibility that some of the proposed 3C\textsuperscript{pro} substrates in SVV-001 are better substrates for fusion proteins like 3CD than for mature isolated 3C\textsuperscript{pro}. A final caveat is that the purified SVV-001 3C\textsuperscript{pro} used in these studies is a fusion protein with MBP. Although it is possible that the presence of MBP may affect the cleavage specificity or kinetics of the protease, this is likely to be a minor effect, as other viral proteases have been purified as fusion proteins and substituted for their normal counterparts in a similar fashion showing no significant change in specificity or activity (126). The use of this construct was required for
these investigations, as the fusion protein, but not isolated SVV-001 3C\textsuperscript{pro}, could be maintained in solution after proteolytic removal of MBP with TEV protease \((115, 117)\). Recognition of the L/VP4.1 FP substrate and FQ peptide by the native SVV-001 3C\textsuperscript{pro} during a lytic viral infection confirms the feasibility of using the recombinant SVV-001 3C\textsuperscript{pro} in initial substrate optimization.

While we were successful at developing a highly optimized protease substrate, it is important to note that the rational approach used for optimization did not interrogate all potential substrates of SVV 3C\textsuperscript{pro}. Approaches for protease substrate identification have been developed that are capable of screening a large sequence space for substrates with efficient turnover \((120, 127)\). L/VP4.1 has the potential to be optimized further using one or a combination of these approaches to test for extended substrates or to test for all possible combinations of P1-P4 residues, potentially leading to L/VP4.1 variants with improved properties.

An ideal substrate for a peptide prodrug is one that is not cleaved by any other protease found in the human body. Unlike the substrates usually conjugated in other peptide prodrugs, the L/VP4.1 substrate described here is a viral polyprotein sequence cleaved by a viral protease. Because neither the protease nor substrate is found in uninfected human cells, this approach substantially lowers the chance of non-specific recognition and activation by cellular proteases. To confirm our hypothesis, we performed an \textit{in silico} search of the MEROPS database which contains cleavage data on over 3,000 proteases including human and human pathogens \((128)\). No human proteases were predicted to cleave L/VP4.1. A single hit for Southampton virus, a calicivirus that causes acute gastroenteritis was also noted \((129)\). Therefore, any peptide prodrug which utilizes a viral protease and substrate pair, such as an
L/VP4.1 based prodrug, is likely to have superior selectivity relative to prodrugs activated by endogenous cellular proteases.

A prodrug based on the L/VP4.1 substrate introduces a novel form of VDEPT based on a wild-type virus. This strategy combines multiple advantages of the antibody-directed enzyme prodrug therapy (ADEPT) and VDEPT concepts while overcoming a few of the limitations. No genetic engineering of the virus is needed because the wild type virus already contains active protease and specifically homes to the tumor site. With this approach, the virus plays a more important role than solely as a transport or targeting vector for the enzyme gene; this strategy incorporates the intrinsic ability of the virus to lyse cancer cells in combination with the peptide-prodrug to increase the therapeutic effect on the tumor. The existence of both an intracellular and extracellular pool of SVV 3Cpro during a cellular infection allows for increased flexibility when designing the peptide prodrug (130).

Although SVV-001 has been shown to be effective in vitro and in patient derived xenografts of variant SCLC, high viral titers are needed to eradicate tumors in vivo (119). Our proposed approach to improving the efficacy of SVV-001 is to combine virotherapy with a peptide prodrug of a potent cytotoxic agent, activated by SVV-001 3Cpro. A prodrug of this kind is expected to be highly selective because of the high specificity of the virus as well as the uniqueness of the SVV-001 3Cpro and its substrates. Such a VDEPT strategy may also allow for a “bystander effect” where uninfected tumor cells could be killed by the active cytotoxic agent released by neighboring infected cells (131, 132). An added benefit of this novel combination is the increased effect on tumor cells before neutralizing antibodies are produced against the virus and render it inactive. A novel combination of this kind may prove to be a highly effective therapy for multiple SVV-permissive cancer types. Finally, these
data represent a proof of concept for other oncolytic viruses with highly active and sequence specific proteases.

This study characterized the substrate requirements of the SVV 3C^pro and identify an optimized substrate with high turnover efficiency for incorporation into a protease activated peptide prodrug. The substrate studies performed here have identified a lead substrate efficiently cleaved by the SVV-001 3C^pro, which may be used as the basis for a peptide prodrug amenable to combination with SVV virotherapy as a novel form of VDEPT.
Chapter 2. Design, execution, and analysis of pooled in vitro CRISPR/Cas9 screens: Identifying essential genes for SVV

History of Genome Wide Screens

The elucidation of gene function is important for further understanding of cellular biology. Both loss of function (LoF) and gain of function (GoF) screens can be used to identify novel protein functions by altering gene expression. While GoF experiments have been possible in mammalian systems for some time using genomic, cDNA library screening and transposon insertion gene trapping, the diploid nature of the mammalian genome has made LoF screening more challenging.

Historically, *Saccharomyces cerevisiae*, a species of yeast, was used as the model system of choice to dissect eukaryotic biological processes due to fast generation time, ease of manipulation in a laboratory setting, and the fact that it can grow in a haploid state. Forward genetic screens for developmental genes in multicellular eukaryotes were most famously performed in the fruit fly *Drosophila melanogaster* by random mutagenesis of the haploid gametes of the parental generation. These model systems were critical to the discovery and understanding of many mammalian biological processes.

The discovery of RNA interference (RNAi) allowed LoF screens to be performed in cell lines derived from mammalian systems representing diverse normal and cancerous backgrounds with interesting genotypic and phenotypic features (133, 134). Despite the power of RNAi, which continues to be used to elucidate many diverse biological processes, incomplete knockdown – even when potent shRNAs are used – can result in a lack of necessary sensitivity as well as off-target effects can complicate identification of validated hits (135-
The discovery and development of two related near-haploid human cell lines, KBM7 and HAP1, allowed for the first time genome-wide insertional mutagenesis screens (138, 139). However, gene trap mutagenesis is only feasible in haploid cell lines, limiting the broad applicability of the method. Although powerful and successful at discovering genes involved in many biological processes, each of these technologies has particular limitations. Recently, the CRISPR/Cas9 technology has allowed for genome wide screening of biological processes in diploid mammalian systems.
The CRISPR/Cas System

Similar to RNA inference pathway utilized by eukaryotic cells to defend against foreign genetic material, a majority of prokaryotic cells exploit the CRISPR/Cas system to defend against foreign genetic material from bacteriophages, viruses, and plasmid DNA (140-143). Within the genomes and plasmids of some prokaryotic cells, loci of clustered regularly interspaced short palindromic repeats (CRISPR) have been identified separated by spacers of variable sequences. Sequencing of the spacers in specific strains of cells, like *Streptococcus thermophilus*, found high sequence homology to DNA sequences found in bacteriophages to which that particular bacterial strain is resistant. Adjacent to the CRISPR loci are multiple genes encoding enzymatic proteins known as Cas (CRISPR-associated) genes. Studies further showed that the spacers within the CRISPR loci confer resistance to bacteriophages, viruses, and plasmid DNA to which they are homologous. Moreover, the CRISPR loci are rapidly altered and evolve as cells are exposed to new foreign genetic material, suggesting the CRISPR/Cas system is a method of adaptive immunity within prokaryotic cells (140, 141).

Mechanistically, the CRISPR loci are transcribed and processed into small CRISPR RNAs (crRNA) by specific Cas proteins. Certain Cas proteins, after interaction with the crRNA, are directed to the invading genetic material and bind to the recognized sequence found within the crRNA, ultimately leading to the degradation of the invading genetic material. The proto-spacer, the DNA/RNA sequence in the “invading” genetic material that is contained in the CRISPR loci spacers, is only recognized and cleaved when a proto-spacer adjacent motif (PAM) sequence is present directly downstream (141, 143). Studies of multiple CRISPR/Cas
systems in prokaryotic cells determined three main types of CRISPR/cas systems. Type II systems are unique in that these systems utilize a trans-activating crRNA (tracrRNA) that recognize the repeated sequences in the immature crRNA to activate processing to the mature crRNA in the presence of the Cas9 protein (144). These studies identified Cas9 as a dual-RNA guided DNA endonuclease capable of causing double stranded breaks in foreign DNA with proto-spacer sequences homologous to the crRNA spacer sequence. The seminal study describing the function of the Cas9 protein from *Streptococcus pyrogenes* further found a recombinant chimeric RNA molecule containing essential features of the crRNA and tracrRNA could also target the Cas9 protein to specific DNA sequences (145). Additionally, it was possible to program the RNA-guided DNA endonuclease activity of Cas9 using the chimeric single guide RNA (sgRNA) to specifically cleave targeted DNA sequences. The discovery of a programmable RNA-guided DNA nuclease quickly lead to the development of the CRISPR/Cas9 system as a specific and efficient way to target and edit genomic DNA (145, 146).

In the context of eukaryotic cells, the CRISPR/Cas9 system uses the DNA nuclease, Cas9, to cause double stranded DNA breaks at a specified locus in the genome using single guide RNAs (sgRNA) complementary to the target DNA sequence. When repaired, small insertion or deletions can be incorporated into the DNA sequence which lead to frame shifts that result in nonsense translation, introduction of premature stop codons, and effectively knock out targeted genes (145, 147, 148). Alternatively, a catalytically dead Cas9 can be utilized as a DNA binding protein along with transcriptional activators or repressors to activate or repress gene expression, respectively.
Types of Human Genome Wide CRISPR Libraries

Currently, there are libraries designed to cause three types of gene expression modification: knockout, repression (CRISPRi), and activation (CRISPRa). Based on the biological question being investigated, the type of library to use may be obvious. For example, the identification of certain proteins, such as viral receptors, may require complete absence of the protein, and therefore complete gene knockout, to generate profoundly resistant cells. In this case, a knockout library would most likely have the greatest probability of identifying that receptor. For loss of function screens, it is recommended to target early, constitutively expressed exons – especially those predicted to result in a non-functional or dominant negative protein. In other instances, gene repression rather than knockout may be more suitable due to more homogeneous repression in cell lines with complex karyotypes. The CRISPR libraries described in this chapter utilize the Cas9 nuclease derived from *Streptococcus pyogenes* as part of a pooled screening approach (147, 148). Cas9 nucleases from other organisms have been characterized, but are not compatible with the existing CRISPR libraries due to differences in PAM sequences, and therefore, will not be discussed. CRISPR screens can additionally be performed using arrayed sgRNAs, testing single or a small number of sgRNA(s) per well in one or more multi-well plates. Although arrayed libraries can be useful and highly successful, the focus of this chapter will only be on pooled screens using sgRNA libraries consisting of many thousand elements. Testing of tandem sgRNAs, or combining multiple sgRNAs in a single cell, in a single screen to elucidate pathways or synthetic lethality relationships will not be discussed in this chapter.

*CRISPR knockout libraries*
The most common CRISPR sgRNA libraries used for \textit{in vitro} screening are the genome-wide knockout libraries generated by the Sabatini/Lander lab and the Zhang lab. Initially described in Shalem et al. and Wang et al., the most recent version of each sgRNA knockout libraries target over 18,000 genes in the human genome. The genome wide CRISPR knockout library (GeCKO) from the Zhang lab also contains sgRNAs targeting over 1,000 micro-RNAs (miRNA) (149). Each group created algorithms to determine sgRNA target sequences within the genome for each gene, focusing specifically on 5’ constitutively expressed exons. The Sabatini/Lander lab library additionally incorporated sgRNA design rules to increase cleavage efficiency and decrease off-target effects (150, 151). Both libraries have the advantage of being commercially available as pooled plasmid libraries. The Zhang lab library has the option of a 1-plasmid or 2-plasmid system. The 2-plasmid system has the advantage of higher sgRNA library titers and ability to express Cas9 from different constructs (149, 152). Additionally, the 2-vector system allows for the generation and selection of an individual Cas9-expressing cell line clone, which removes some inherent variability from cell to cell compared to the 1-vector system. However, the 2-vector system can require a longer generation time than the 1-vector due to multiple selection steps. The Sabatini/Lander lab library is available in a 1-plasmid system (150, 151).

Recently a second-generation CRISPR knockout library was described by Hart et al. (153). The Toronto KnockOut (TKO) library was designed to only utilize effective targeting sgRNAs with 0-1 genomic off-target sites. This library is now commercially available as two half libraries, similar to the first generation CRISPR libraries. Also of note, a very recent publication by Doench et al. described a new CRISPR sgRNA library that incorporated more detailed sgRNA design rules to maximize cleavage efficiency by Cas9 while decreasing off-
target effects (154). This optimized library is now available commercially and is the first of the next wave of genome-wide knockout libraries.

The Sabatini/Lander lab has additionally designed subpool libraries targeting genes with related functions, such as kinases or cell cycle proteins (150, 151). If a screen is intended to explore a narrow hypothesis, such as to identify kinases that are required for survival in the presence of a small molecule inhibitor, investigators may prefer to use a focused sgRNA library rather than a genome-wide approach. Focused libraries have the advantage of being smaller, allowing for the use of fewer cells and less reagents. The Sabatini/Lander libraries employ a Cas9 that is under the control of a doxycycline inducible promoter. The temporal regulation of Cas9 expression allows for an additional control of genome modification and initiation of the knockout screen (150). Although this chapter will not cover these in detail, other CRISPR knockout libraries have been described including ones which specifically target functional protein domains within genes, instead of just 5’ exons, to increase the probability of identifying indels causing deleterious effects to not only protein expression but also protein activity (155).

With any of the gene knockout libraries, screens can be designed for either positive or negative selection of essential sgRNAs (Figure 11). The nomenclature of the screens are similar to those used to describe RNAi screens (156). Positive selection screens identify genes needed for a cell to be sensitive to the selection mechanism applied during the screen. In this case, most cells will die from the selection and the cells that grow out of the screen contain sgRNAs targeting genes essential for the agent being tested. An example of a positive screen is a CRISPR screen seeking to identify genes essential for a lytic virus, where the loss of essential genes will create cells that are resistant to a viral infection. Conversely, a
negative selection or “dropout” screen will identify genes essential for resistance to the selection agent where gene knockout causes these cells to die. For a negative selection screen, most cells will survive the selection and the goal is to identify which sgRNAs are no longer present or are reduced in representation compared to controls. Although both types of screens are capable of identifying true positive hits, negative selection screens can be much more technically challenging as rigorous control samples and deep sequencing are needed to identify statistically significant changes in sgRNA representation when the majority of cells survive the screen. Statistical considerations will be discussed in a following section. Additionally, screens can be performed using methods of enrichment other than cell growth or viability. For example, screens can be performed using reporter-based assays and enrichment by fluorescence-activated cell sorting (FACS). The sensitivity of FACS sorting and ability to sort using multiple markers can yield powerful results, but multiple rounds of enrichment may be required (157).
Figure 11. Types of CRISPR screens. Once a selected cell line has been transduced with the necessary components of a CRISPR library (top panel), two types of screens can be performed: positive selection screens and negative selection screens. In positive selection screens, cells that undergo Cas9-dependent gene manipulation/modification that results in resistance to or increased survival under the specific screen selection pressure (i.e. therapeutic drug or virus) will be identified as all other cells preferentially succumb to the selection agent (middle panel). Conversely, in negative selection screens (bottom panel), most cells will survive the selection agent and only those where the targeted genetic manipulation/modification results in increased sensitivity to or loss of survival under the screen selection pressure will be lost during the screen.
CRISPRi libraries

Alternative options to knockout libraries are the more recently described CRISPR-based interference (CRISPRi) libraries. In contrast to the gene knockout libraries, CRISPRi libraries utilize a catalytically inactive Cas9 (dCas9) to repress transcription. dCas9 protein, targeted to specific DNA sequences by sgRNA, can inhibit transcription by directly blocking the RNA polymerase or by repressing transcription as a fusion protein to a repressive effector domain ([158]). The Weissman lab and others have previously shown that repressive chromatin modifier domains, like Krüppel-associated box (KRAB), fused to the dCas9 protein improve transcriptional silencing compared to the dCas9 protein alone. This CRISPRi library contains sgRNAs targeting approximately 16,000 human genes for transcriptional silencing by a dCas9-KRAB fusion protein as a 2-plasmid system. This CRISPRi library is currently commercially available, and other CRISPRi libraries may follow ([159]).

CRISPRi allows for a different type of genomic modification than the knockout libraries. Because the dCas9 is not catalytically active, CRISPRi does not lead to permanent DNA alteration but instead represses transcription from the promoter to which dCas9 is recruited. The level of transcriptional repression, however, can vary depending on many factors such as sgRNA and targeted DNA sequence ([158]). In many cases, CRISPRi will not be as efficient as gene knockout and instead may have an effect on expressed protein levels similar to that of RNAi and may thus be a preferred approach in cases where complete protein knockout is not required. CRISPRi is reversible if an inducible dCas9 is utilized, which allows for the additional variable of turning on/off transcription repression to be included in screen design ([159]).
CRISPRa libraries

A slightly different approach in gene candidate identification can be employed by performing a GoF screen using CRISPR gene activation (CRISPRa) libraries. These libraries exploit a similar method of transcriptional regulation as the CRISPRi libraries, fusing the dCas9 protein to known activation domain proteins (160, 161). However, the goal of CRISPRa libraries is to increase gene expression levels so screens using these libraries, unlike the gene knockout and CRISPRi library screens, will be GoF screens.

There are two main CRISPRa libraries that are currently commercially available to researchers. Both libraries utilize the herpes virus tetrameric transcriptional activation domain VP64 to assist in activating gene expression of targeted coding sequences, but differ in terms of other modifications to the sgRNAs and dCas9 protein. Weissman and colleagues developed a dCas9 fusion protein based on the SUperNova tag (SunTag) system for signal amplification (162). The dCas9 fusion protein contains a C-terminal tail of 10x GCN4 peptide epitopes, which are recognized by single-chain variable fragment (scFv) antibodies fused to the VP64 activation domains. The SunTag system recruits multiple copies of the VP64 domain to a single dCas9 protein, amplifying the transcriptional activity of the targeted gene (159). In contrast, the CRISPRa library designed by Zhang and colleagues use a dCas9-VP64 fusion protein combined with a modified sgRNA and a trimeric fusion protein complex to make the Synergistic Activation Mediator (SAM) library. The modified sgRNA sequence contains two RNA hairpin aptamers recognized by the bacteriophage coat protein MS2. The trimeric fusion protein is made up of the MS2 protein for binding the sgRNA and two transcription activation domains from NF-κB (p65) and heat shock factor 1 (HSF1). The
combination of VP64, p65, and HSF1 activation domains synergistically increases gene expression of the targeted coding sequences (163).

Both CRISPRa systems have a capability to increase gene expression up to 40-fold above baseline levels, but the transcriptional induction will vary depending on the targeted gene (159, 163). These libraries give researchers the unique advantage of performing genome wide GoF screens. However, the addition of a third component for maximal transcriptional activation, either the scFc-VP64 or MS2-p64-HSF1 fusion protein, could add complexity to efficient transductions and expression depending on the cell line used for the screen.
Choice of Cell Line for Pooled CRISPR Screening

There are many considerations that may influence the optimal choice of a cell line for a genome wide screen. In some cases, it may be best to perform the screen in multiple cell lines to avoid specific caveats such as the genetic background of the cell line or transduction efficiency. The efficiency of any type of genetic manipulation, whether reversible or irreversible, will depend on the number of target genetic loci that need to be modified. Most normal human cells are diploid and therefore, will contain two copies of a gene. However, many cancer cell lines have aberrant numbers of chromosomes and focal gene amplification due to inherent genomic instability (164). If cell line A has 8 copies of a given gene and the cell line B has 2 copies, by probability, complete gene knockout will be more probable in a given cell from line B than from A. Therefore, the exact same screen performed in a diploid cell line may yield more robust data with higher signal to noise ratio than a screen performed in a hyperdiploid cell line due to the increased difficulty of modifying every loci in the polyploid cell line. Similarly, a screen performed in a haploid cell line may be more sensitive and yield higher quality data than a screen performed in a diploid or hyperdiploid cell line. For this reason, a number of successful LOF screens using gene trap mutagenesis were performed in haploid cell lines due to the ease of creating complete gene knockouts (138, 139). Although the copy number variation for every single gene in a given cell line may not be known, the karyotype of many cell lines is available from the commercial source they are purchased from, or can be obtained from karyotyping services at many academic core facilities. Karyotype data can be used to infer average gene copy number and allow for the researcher to decide upon a cell line.
An additional caveat for consideration specifically for CRISPR knockout screens is the status of DNA repair pathways in a specific cell line. In CRISPR knockout screens, the Cas9 nuclease creates DNA double strand breaks (DSBs) in the sequence targeted by the sgRNA (reviewed in (165)). The DSB is then repaired by one of two DNA repair pathways: homology directed repair (HDR) or non-homologous end joining (NHEJ). Where HDR can repair the break precisely using a homologous DNA donor like a sister chromatid, NHEJ promiscuously trims and re-ligates the DSB leading to small insertions or deletions (indels). These indels in a coding sequence can ultimately cause frameshifts and premature stop codons upon protein translation, abrogating protein expression. Gene inactivation is dependent on NHEJ. Therefore, if both pathways are functional within a cell there is a lower probability of complete gene knockout at every locus, as both pathways can repair the DSB. However, cells lines where HDR is partially or completely defective, as is the case with many cancer cell lines, DSB repair is dependent on NHEJ (166). Transducing cell lines defective in HDR with the CRISPR knockout libraries may have better success rates in generating complete gene knockout cells and ultimately generate more dependable “hits” during a screen. Investigators should empirically test the efficiency of gene knockout via CRISPR/Cas9 across a panel of possible cell lines of interest to determine which cell lines may be most suitable. Furthermore, expression of Cas9 and additional confirmation of DNA modification should be confirmed in each line to validate the efficiency of the CRISPR/Cas9 in a given cell line prior to the start of any screen. For dropout screens in particular, it may be necessary to derive a single cell clone from a pool of Cas9 expressing cells with high expression of Cas9 and confirm gene editing efficiency using an sgRNA targeting an essential gene.
The level of susceptibility of a cell line to the selection agent in the screen may influence the number of gene hits obtained and overall success of a screen. Across a panel of cell lines treated with a selection agent, there will be a range of sensitivity. Exquisitely sensitive cell lines will not need high levels of the selection agent to show a loss of cell viability or change in doubling time. However, only absolutely essential genes may be identified from the screen because of the high stringency of the selection. More resistant cell lines will need higher levels of the selection agent and may thus may be capable of identifying genes with more subtle phenotypes. There may also be more cells that survive or grow out from the screen in this case but high throughput sequencing and secondary screens will confirm the important genes for the selection agent. The choice of a cell line may dictate what genes and how many are identified from the screen.

Lastly, the transduction efficiency of a cell line will be very important for the success of the screen as the ability to infect the cell line with one or multiple lentiviruses is necessary for the success of the screen. Choosing a cell line that has low transduction efficiency will lead to increased difficulties in generating a pool of cells with acceptable representation of the sgRNA library. Lentiviral transduction of sgRNA libraries is typically performed at MOIs between 0.4 and 0.6 transduction units/cell to ensure each cell contains a single sgRNA (149, 150). The amount of virus needed to achieve these MOIs is dependent on both the titer of the virus and the transduction efficiency of the cell line being tested.
Analysis of CRISPR Screens

Almost every screen described to date has used PCR amplification to prepare samples for sequencing analysis. This method allows for specific amplification of the sgRNAs contained within the surviving cell population and sensitive identification of changes in sgRNA representation between samples. In order to quantitate the proportion of each sgRNA in a given population, it is necessary to design PCR primers to amplify only the portion of the lentiviral backbone containing the sgRNA. Next generation sequencing technology is designed to sequence libraries of nucleic acids that have very high diversity. Therefore, it is critical that the sequencing primers be designed using a staggered approach in order to introduce apparent library diversity while maintaining sgRNA complexity. This is especially important in selected populations that may be dominated by a small number of sgRNAs. This staggered approach should also be taken into consideration when hits are being deconvoluted.

At the conclusion of a typical pooled CRISPR/Cas9 screen, frozen cell pellets will have been collected at different time points throughout the course of the screen that should include at a minimum baseline, experimental, and control cell populations, each with a sufficient number of cells to represent the baseline mutagenized population diversity. DNA can be extracted using standard procedures, however, it is important to make sure that extractions are scaled appropriately so as not to overload the chosen purification approach and unintentionally reduce sample diversity. Since library diversity is biased in the representation of each sgRNA, the population of cells to be screened and analyzed should be sufficiently large to ensure capture of under-represented sgRNAs. For a high quality library, 300-1,000-fold
representation is typical. This representation should be maintained throughout the screen and during library preparation. To reach and maintain the suggested fold representation during a screen involves multiple variables dependent on the cell line being utilized and the screen itself. For most screens performed to date, however, approximately 100-200 million cells are transduced and subsequently screened (149, 150).

Level of Sequencing Depth

After library preparation, a further consideration is sequencing depth. The necessary sequencing depth required to achieve adequate interrogation of the library complexity will be different for each CRISPR library and screen. Based on data from our screens as well as others, $1-2 \times 10^7$ reads is more than sufficient to adequately sequence a complex library of $\sim 1 \times 10^5$ elements at the plasmid or baseline infected cell population level. In the case of positive selection screen with strong selective pressure, useful results can be obtained with only a few million reads. However, negative selection screens, where most cells survive and changes in representation may be subtle, may require much deeper sequencing up to $1 \times 10^8$ reads.

To avoid guess-work and economize sequencing resources, it may be preferred to determine empirically the appropriate number of reads to adequately cover the library of interest. Next generation sequencing is a population-sampling problem that is not unlike the enumeration of different species in ecological surveys. Techniques developed for ecology can be used to determine the population complexity of a sequencing reaction based on a small sampling (167). Preseq is one implementation of this approach that can make accurate predictions of library complexity from fewer than $1 \times 10^6$ reads (168). It may thus be preferable to
extrapolate the appropriate target read depth in this manner, especially if multiplexing is available (Figure 12A).

**Quality Control and Normalization Methods**

After library sequencing, the researcher may receive FASTQ files containing the resulting reads from the sequencing reactions. High quality sequencing data is the foundation of the analysis and therefore requires careful consideration of downstream processing. Standard next-generation sequencing QC measures should be used to determine base quality within reads. Resulting reads should be trimmed to eliminate poor quality bases, but in practice it may be sufficient to proceed directly to sgRNA enumeration. No normalization is required; however, since the sequenced libraries may have different numbers of reads, it is important to control for library size in downstream analyses. For figures, it may be appropriate to express data in terms of log2 counts per million reads. Library size factors should be included in any statistical models used.
Figure 12. Example data analysis of a genome-wide pooled CRISPR screen. Shalem et al. screened for sgRNAs targeted genes that allow survival of BRAF mutant A375 cells in the presence of vemurafenib (PLX) using their GeCKO library. **A.** Extrapolation of distinct reads as a function of total reads from a sample of $1 \times 10^6$ reads using Preseq. Grey shaded area represents 95% CI. **B.** Empirical cumulative distribution of cells treated with DMSO or PLX at either 7d or 14d with duplicates demonstrates loss of library diversity after selection with a BRAF inhibitor. **C.** A waterfall plot of the most highly enriched sgRNAs shows the marked enrichment of $NF2$ and $CUL3$ in PLX-treated cells. **D.** The positive enrichment of $NF2$ and $CUL3$ is even more evident when fold change is plotted as a function of average abundance. Data in this figure were reanalyzed from reference 149.
The sgRNA representation of each sequenced library can be compared to that of the initial plasmid library. This allows confirmation that all sgRNAs were represented at the initial condition and determination of the degree to which selection occurred in the course of the screen. One can visualize the degree to which multiple libraries have similar representation is by examining a plot of the cumulative distribution as a function of normalized reads. Shifts of this curve to the left represent loss of diversity from the initial plasmid library (Figure 12B). Another way to approach this is with a waterfall plot, which displays the representation of each sgRNA relative to the median abundance in the library (Figure 12C). This can be a quick way to visualize the proportion of sgRNAs that are relatively over or under-represented as well as to show how individual sgRNAs move in this distribution under different conditions.

Type of Statistical Analysis

Once the representation of individual sgRNAs has been determined for each sample, the statistical significance of the changes in sgRNA representation can be determined. Statistical analysis of screen results is dependent on some specific properties of pooled screens: screen data is represented as count data, the variance of guide representation is not linearly correlated to the level of representation, and finally, these data do not follow a Poisson distribution. Fortunately, these characteristics are similar to the statistical challenges that have arisen in the analysis of differential gene expression from RNA-seq data. Sophisticated algorithms have been developed in order to estimate dispersions and model mean/variance trends within the data such as LIMMA and edgeR (169, 170). These allow weighting of tag-wise and gene-wise results if biological replicates are available.
It is generally preferable to use biological replicates for pooled library screening. However, it is sometimes possible to obtain useful results without true biological replicates if the screen in question results in a small number of hits and is of very high stringency. Biological replicates are typically used to model the dispersions within a group of replicates. However, it is possible to model common dispersions between control and experimental conditions. This approach may fail to converge if the experimental library has much lower population complexity after selection. When biological replicates are available, it is possible to determine accurate \( p \)-values for individual sgRNAs (tag-wise) as well as individual genes (gene-wise) (Figure 12D).

Tag-wise \( p \)-values can be determined using most approaches designed for RNA-seq differential expression analysis as well as turnkey approaches such as RIGER \((171)\), RSA \((172)\), HitSelect \((173)\), and MAGeCK \((174)\). Each of these software packages uses slightly different algorithms for hit selection and it may be necessary to inspect the results of nominated hits graphically with respect to the relationship to other targeting sgRNAs. If non-targeting sgRNAs are present in the library, they can be especially useful in delineating the null distribution and setting empirically determined cutoffs. The approaches mentioned here typically require some programming experience; however, they tend to offer the greatest flexibility. The edgeR package offers a simple 2-class test, exactTest() for basic implementations as well as support for generalized linear models that allow for more complicated designs.

Gene-wise \( p \)-values can be more challenging to calculate; however, MAGeCK has built-in support for ranking by gene. Ranking by gene is again not unlike gene set enrichment analysis in which the researcher would like to identify gene sets whose members are
significantly up or down-regulated. In this case, each gene set is comprised of the sgRNAs that target the gene of interest. Multiple approaches have been designed for this purpose including rotation based approaches such as ROAST (175), and competitive gene set tests such as Camera (176). It should be noted that although fast, Camera assumes inter-gene correlation, which is not present between sgRNAs in a pooled screen. Thus, inter-gene correlation should be manually set to something low such as .01.
Confirmation of CRISPR Screen Hits

Once the sequencing data from a screen has been analyzed, it is necessary to validate each hit identified in an assay similar to the screen prior to other downstream steps. This orthogonal screen should separate true positive hits from false positive hits, although the number of false positive hits will vary from screen to screen. Validated hits should be further examined in a secondary screen. This screen will be dependent on the screen selection reagent and list of possible hits. For example, if a primary screen is used to identify sgRNAs enriched only in the presence of the allosteric mTOR inhibitor rapamycin, an appropriate secondary screen may be involve creating cell lines with protein knockdown of possible hits, measuring proliferation by mitochondrial conversion of a colorimetric reagent, and testing resulting sensitivity/resistance to rapamycin and/or an mTOR kinase inhibitor. To create these lines, several methods are available, such as using the support of parallel systems (e.g., si and shRNA knockdown experiments) or re-creating individual gene knockouts using the same sgRNAs identified in the screen. Conversely, for CRISPRa screens, cDNA overexpression may be used to mimic the transcriptional activation of a gene observed during the screen. Once hits have been validated using a secondary screen, additional experiments should be performed to validate individual genetic and/or transcription level modifications. Depending on the type of screen performed, every type of validation described here may not be necessary.

Confirmation of Cas9-dependent gene modification

For CRISPR gene knockout screens, the first validation experiment should confirm that the sgRNA targeted gene has undergone a modification in the DNA sequence created by the
Cas9-dependent double strand break and subsequent repair. Mis-paired heteroduplex DNA substrates can be identified via treatment of the cell pools with endonucleases prior to detection using gel electrophoresis. Deployment of the Surveyor assay, using Cel1 to detect deletions and T7E1 to call attention to single nucleotide modifications, is in wide use (177, 178). Another way in which to identify small-sized indels in heterogeneous populations of cells is to run a high resolution melt (HRM) analysis in which PCR amplicons of 70-350 bp drive the differences in melting curves when CRISPR-generated indels have been created (179). A less common, pricier approach would be to use fragment analysis via automated capillary electrophoresis (ACE), in which fluorescent tags are applied to the amplicons prior to running the samples on a capillary electrophoresis instrument. Subsequent analysis of these fragment lengths can reveal even single nucleotide changes. If read lengths >10 kb are desired, as in large scale genomic inversion/deletion assays, a recommendation would be to use single molecule real time sequencing (SMRT) as described in (180).

**Confirmation of transcript level changes**

Especially important for CRISPRa and CRISPRi screens, in which transcription level changes are desired, is to confirm that such changes occur using the CRISPRa/i methods. Real-time quantitative PCR (RT-qPCR) can enable the measurement of transcript levels in both coding and non-coding targets. Sanger sequencing of cDNA in haploid cell lines or NGS of diploid cell lines can examine specific loci. Control experiments using non-targeting sgRNAs should be performed in parallel to confirm that transcription level changes are due to Cas9-dependent gene activation/repression.

**Confirmation of protein level changes**
For most phenotypic changes, confirmation of protein level changes will be the most important validation step. This will be useful for all types of CRISPR screens as it is usually the protein translated from the gene of interest that leads to downstream signaling and phenotypes. Western blots are often the simplest method used to provide data regarding protein translation. However, specific antibodies are needed and depending on the target, it may be difficult to find ones of good quality that recognize a portion of the protein 5’ to the sgRNA binding site. The use of downstream gene ontology/biological pathway analyses can complement target validation and provide biological context to hits.

*Further Experiments*

Once the necessary validation experiments have been performed to confirm that the CRISPR screen created modifications in the specific gene/mRNA/protein being targeted, further experiments should be performed to identify and characterize the significance of the hit in terms of the selection agent tested in the screen. For example, if the screen was performed to identify proteins essential for a virus to enter and replicate within a cell, experiments should attempt to characterize the role the identified protein may play in the viral life cycle. These experiments may involve performing GoF experiments for CRISPRi and CRISPR knockout hits or loss-of-function experiments for CRISPRa hits. Rescue experiments, to add back the missing protein via cDNA expression or to actually repair the genomic defect using CRISPR-Cas9, are suggested options. Similarly extensions into models of higher complexity may be important for particular phenotypes.
Study Introduction

Multiple oncolytic viruses are now in active clinical development as cancer therapeutics (39, 181). As mentioned previously, a modified herpes simplex virus recently became the first oncolytic virus FDA-approved for use in human cancer patients (40). Relative to other oncolytic viruses in clinical evaluation, SVV is notable for its small size, exceptionally rapid doubling time, high selectivity for neuroendocrine cancer cells, and the absence of preexisting neutralizing antibodies in patients(59). Previous studies in multiple pre-clinical mouse models and early phase clinical trials confirmed the safety and potential efficacy of SVV as a novel cancer treatment but clinical development of SVV has been hampered by a lack of understanding of cellular determinants of infection, including identification of the cellular receptor for SVV (20, 42, 60, 63-65). Given that even among high-grade neuroendocrine tumors only a subset of cell lines and tumors support SVV replication, there is a clear need for a biomarker for SVV permissivity (20, 59, 62).

Exploiting the power of genome wide CRISPR knockout (GeCKO) screens, we performed pooled genome wide loss-of-function screens to identify essential proteins for an SVV infection. Parallel screens were completed in two SVV permissive cell lines- the SCLC cell line, H446 and the CML haploid cell line, HAP1. The results from the screens were analyzed using divergent methods as the level of stringency varied between the screens. The GeCKO screens were successful in the discovery of important proteins in SVV permissivity. One of the proteins, anthrax toxin receptor 1 (ANTXR1) was found in both screens. The results in this chapter describe the GeCKO screens and the discovery of ANTXR1 as the cellular receptor of SVV. As discussed in the previous section, further experiments to confirm and
characterize the ANTXR1 as the cellular receptor for SVV were also completed and are presented in chapter 3.
Materials and Methods

Reagents and Bacterial Strains

Polymerase chain reactions (PCRs) were carried out using a GeneAmp PCR System 9700 Thermocycler (Applied Biosystems). PCR fragments were purified using QIAquick PCR Purification Kit (Qiagen). Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen). Competent DH10B and Stbl3 cells were purchased from Invitrogen. Plasmids were isolated and purified from bacteria using QIAquick Spin Miniprep Kit (Qiagen). Sanger sequencing for individual clones and plasmids was performed by Genewiz, Inc. Nucleotide and protein sequence alignments were performed in Geneious Pro 4.7.6.

Cell Lines and Viruses

All cell culture media was produced by the Memorial Sloan Kettering Cancer Center (MSKCC) Media Prep core facility. HAP1 cells were purchased from Haplogen GmbH and maintained in Iscove’s Modified Dulbecco’s Medium (IMDM) supplemented with 10% fetal calf serum. All other cell lines described in this chapter were purchased from ATCC. HEK 293T/17 cells were maintained in high glucose Dulbecco’s Modified Eagle Medium (DMEM HG) supplemented with 10% fetal calf serum and 1 mM sodium pyruvate. NCI-H446 were maintained in RPMI 1640 supplemented with 10% fetal calf serum and 10 mM HEPES. All cell lines are routinely confirmed by STR analysis and confirmed mycoplasma negative by DDC Medical. SVV and SVV-GFP was cultured, purified, and titered as previously described (59, 62).

Human GeCKO v2 Library
The Human GeCKO v2 library was obtained as two half libraries (Library A and B) in the lentiGuide-Puro plasmid backbone (Addgene plasmid #52962) as a gift from Feng Zhang. The MSKCC RNAi core facility amplified the pooled libraries by electroporation of Endura electrocompetent cells (Lucigen) as described previously (149, 152). Plasmid DNA libraries were used as a template in a nested PCR to first amplify the section of the plasmid containing the sgRNA (Table 3; primer #1-2) and subsequently to add Illumina sequencing adaptors and barcodes (primer #3-18). The nested PCR products were then sequenced for confirmation of sgRNA representation using an Illumina HiSeq2500 (primer #19).
<table>
<thead>
<tr>
<th>Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1- lentiCRISPR amplification F</td>
<td>5’-CTTTAGTTTGTATGTCTGTCTATTATGTCTACTATTCTTTCC-3’</td>
</tr>
<tr>
<td>2- lentiCRISPR amplification R</td>
<td>5’-AATGGACTATCATATGCTTCACCTGAACTTAA-3’</td>
</tr>
<tr>
<td>3- HiSeq F#1</td>
<td>5’- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTATTCTGTTGGAAGGAGCAAAACCCG-3’</td>
</tr>
<tr>
<td>4-HiSeq F#2</td>
<td>5’- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTATTCTGTTGGAAGGAGCAAAACCCG-3’</td>
</tr>
<tr>
<td>5-HiSeq F#3</td>
<td>5’- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTATTCTGTTGGAAGGAGCAAAACCCG-3’</td>
</tr>
<tr>
<td>6-HiSeq F#4</td>
<td>5’- AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCTATTCTGTTGGAAGGAGCAAAACCCG-3’</td>
</tr>
<tr>
<td>7-HiSeq R#1</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>8-HiSeq R#2</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>9-HiSeq R#3</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>10-HiSeq R#4</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>11-HiSeq R#5</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>12-HiSeq R#6</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>13-HiSeq R#7</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>14-HiSeq R#8</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>15-HiSeq R#9</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>16-HiSeq R#10</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>17-HiSeq R#11</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>18-HiSeq R#12</td>
<td>5’- CAAGCAGAAGACGCATACGAGATCTGATGTGACTGAGAAGGTTACGACG TGIGCCTTTCCGATCTTTCTTACTATTCTTTCCCTCCTGACTGT-3’</td>
</tr>
<tr>
<td>19-HiSeq seq primer</td>
<td>5’-ACACTCTTTCCCTACACGACGCTCTTTCCGATCT-3’</td>
</tr>
<tr>
<td>Number</td>
<td>Name</td>
</tr>
<tr>
<td>--------</td>
<td>--------------</td>
</tr>
<tr>
<td>20</td>
<td>lentiGuide F</td>
</tr>
<tr>
<td>21</td>
<td>lentiGuide R</td>
</tr>
<tr>
<td>22</td>
<td>ANTXR1.1 F</td>
</tr>
<tr>
<td>23</td>
<td>ANTXR1.1 R</td>
</tr>
<tr>
<td>24</td>
<td>ANTXR1.2 F</td>
</tr>
<tr>
<td>25</td>
<td>ANTXR1.2 R</td>
</tr>
<tr>
<td>26</td>
<td>ANTXR1.3 F</td>
</tr>
<tr>
<td>27</td>
<td>ANTXR1.3 R</td>
</tr>
<tr>
<td>28</td>
<td>G3BP1 F</td>
</tr>
<tr>
<td>29</td>
<td>G3BP1 R</td>
</tr>
<tr>
<td>30</td>
<td>KCNJ1 F</td>
</tr>
<tr>
<td>31</td>
<td>KCNJ1 R</td>
</tr>
<tr>
<td>32</td>
<td>LSMEM2 F</td>
</tr>
<tr>
<td>33</td>
<td>LSMEM2 R</td>
</tr>
<tr>
<td>34</td>
<td>NR2C2 F</td>
</tr>
<tr>
<td>35</td>
<td>NR2C2 R</td>
</tr>
<tr>
<td>36</td>
<td>NTHL1 F</td>
</tr>
<tr>
<td>37</td>
<td>NTHL1 R</td>
</tr>
<tr>
<td>38</td>
<td>PLXNB2 F</td>
</tr>
<tr>
<td>39</td>
<td>PLXNB2 R</td>
</tr>
<tr>
<td>40</td>
<td>PPBP F</td>
</tr>
<tr>
<td>41</td>
<td>PPBP R</td>
</tr>
<tr>
<td>42</td>
<td>SCAF8 F</td>
</tr>
<tr>
<td>43</td>
<td>SCAF8 R</td>
</tr>
<tr>
<td>44</td>
<td>TACO1 F</td>
</tr>
<tr>
<td>45</td>
<td>TACO1 R</td>
</tr>
<tr>
<td>46</td>
<td>TEX2.1 F</td>
</tr>
<tr>
<td>47</td>
<td>TEX2.1 R</td>
</tr>
<tr>
<td>48</td>
<td>TEX2.2 F</td>
</tr>
<tr>
<td>49</td>
<td>TEX2.2 R</td>
</tr>
<tr>
<td>50</td>
<td>ZDHHC7 F</td>
</tr>
<tr>
<td>51</td>
<td>ZDHHC7 R</td>
</tr>
<tr>
<td>52</td>
<td>ZNF101 F</td>
</tr>
<tr>
<td>53</td>
<td>ZNF101 R</td>
</tr>
<tr>
<td>54</td>
<td>sgRNA Seq F</td>
</tr>
<tr>
<td>55</td>
<td>ANTXR1 exon2 PCR F</td>
</tr>
<tr>
<td>56</td>
<td>ANTXR1 exon2 PCR R</td>
</tr>
<tr>
<td>57</td>
<td>attB1 cDNA F</td>
</tr>
<tr>
<td>58</td>
<td>attB1 cDNA R</td>
</tr>
</tbody>
</table>

Table 3. Oligonucleotide list for primers used in cloning and sequencing for CRISPR constructs and ANTXR1 expression constructs. Numbers correspond to oligonucleotide identity in the main text. Primer name describes the sgRNA gene target or product the primers will eventually create. Oligonucleotides shown from 5’ to 3’.
**Individual sgRNA Plasmids**

The lentiCRISPR v2 plasmid, which expresses a single sgRNA under the hU6 promoter and the WT Cas9 nuclease under the EFS promoter, was a gift from Feng Zhang (Addgene plasmid #52961). Oligos containing the gene targeting sgRNA with 5’ overhang BsmBI digestion sites were synthesized by Sigma Aldrich (primers #22-53). The oligos were annealed and inserted into the lentiCRISPR v2 backbone as described previously (149, 152). sgRNA sequences and plasmids were confirmed by Sanger sequencing.

**Lentivirus Production**

The lentiviral packaging plasmids pMD2.G (Addgene plasmid #12259) and psPAX2 (Addgene plasmid #12260) were gifts from Didier Trono. Lentiviral plasmids were transfected at a 3:2:1 DNA ratio of lentiviral plasmid:psPAX2:pMD2.G in 1 mg/mL polyethylinimine (PEI; Sigma Aldrich) at a 2:1 PEI:DNA ratio in OptiMEM (Gibco). Media was changed 16 h post transfection. Seventy-two hours post transfection the virus containing media was harvested and filtered through a 0.45 μm PDVF syringe filter (Millipore) to remove cell debris. Lentivirus aliquots were stored at -80°C.

For the human GeCKO library, 20 15 cm² dishes (Corning) were seeded with HEK 293T/17 cells (7.0×10⁶ per plate). The GeCKO libraries A and B were pooled 1:1 (54 μg each library) and co-transfected with psPAX2 (72 μg) and pMD2.G (36 μg) with 432 μL of 1 mg/mL PEI in 36 mL OptiMEM. The transfection mix was then divided equally among the 15 cm² dishes. Sixteen hours post transfection, media on each dish was changed and supplemented with 1 U/mL DNase I (New England Biolabs). The media lentiviral supernatant was
harvested 72 h post transfection and filtered through a 0.45 µm Stericup PVDF filter (Millipore). The virus was then pelleted by ultracentrifugation at 24,000 rpm for 2 h at 4°C. The virus pellet was resuspended in fresh DMEM and incubated overnight at 4°C. Lentivirus aliquots were stored at -80°C.

GeCKO Library Screen

GeCKO lentivirus was thawed on ice and added to OptiMEM supplemented with 32 µg/mL polybrene. After the virus-OptiMEM mix was added to the cells, additional media was added to bring the final polybrene concentration to 8 µg/mL. Media was changed 24 hr after transduction to remove polybrene. Media supplemented with 0.5 µg/mL puromycin (Sigma Aldrich) or 6 µg/mL blasticidin (Fisher Scientific) was changed 48 hr after transduction to select lentiCRISPRv2 or lentiCas9-Blast transduced cells, respectively. The human GeCKO v2 library lentivirus was titered on WT HAP1 and H446 cells as described previously (149, 152). WT cells were transduced with the lentiCas9-Blast lentivirus, allowing for constitutive expression of the DNA nuclease, Cas9. The lentiCas9-Blast plasmid was a gift from Feng Zhang (Addgene plasmid #52962). Transduced cells were selected with 6 µg/mL blasticidin. HAP1-Cas9 cells (2.0×10^8) were seeded equally in 70 15 cm² dishes. H446-Cas9 cells (1.5×10^8) were seeded into 50 15 cm² dishes. GeCKO lentivirus was thawed on ice, mixed in a total volume of 375 mL OptiMEM supplemented with 32 µg/mL polybrene, and divided equally among the HAP1-Cas9 or H446-Cas9 plates. Lentivirus was added at an MOI=0.4 transduction units/cell (TU/cell) for both screens. Additional media was added to each plate to bring the final polybrene concentration to 8 µg/mL. Media was changed 24 hr post transduction to remove polybrene. The media was changed 48 hr post transduction to select
transduced H446-Cas9 or HAP1-Cas9 cells with 0.5 \( \mu \)g/mL or 1.0 \( \mu \)g/mL puromycin, respectively. Transduced cells were allowed to grow for 4 additional days to allow for complete knockdown of sgRNA-targeted genes. On day 7 post transduction in the HAP1 screen, \( 2.0 \times 10^8 \) cells were plated at equal cell density in 40 15 cm\(^2\) dishes and infected with SVV at a MOI=1,000 vp/cell the next day. On day 7 post transduction in the H446 screen, \( 1.5 \times 10^8 \) cells were plated at equal cell density in 50 15 cm\(^2\) dishes and infected with SVV at a MOI=1.0 vp/cell the next day. The remaining cells for each cell line were pooled, pelleted by centrifugation, and stored in -80°C as the corresponding Day 7 post transduction sample. For one week post SVV infection during the HAP1 screen, 15 mL media on the infected plates were exchanged every 3 days to resupply cells with fresh media. Surviving cells were pooled, pelleted by centrifugation, and stored at -80°C as the SVV resistant sample. For two weeks post SVV infection during the H446 screen, 10 mL media on the infected plates were exchanged every 3 days to resupply cells with fresh media. Visible colonies of surviving cells were collected by isolated trypsinization in cloning cylinders and seeded in 1 well of a 24 well plate (Corning). All colonies too small for isolation were harvested by trypsinization and pooled before expansion. Each isolated colony was ultimately expanded from a 24 well to a 75 cm\(^2\) flask as the cells were propagated. Cells from each colony were pelleted by centrifugation and stored in -80°C.

**Identification of sgRNAs**

For the HAP1 screen, extracted genomic DNA from Day 7 post transduction and SVV resistant population was used as a template for the GeCKO v2 library nested PCR and analyzed for sgRNA representation by Illumina HiSeq as described above. Sequenced
sgRNAs were imported from raw FASTQ files, normalized for library size, then converted to log counts per million reads (logCPM). Log fold change was then calculated between control and resistant samples. Based on the distribution of non-targeting sgRNAs, we focused on genes for which ≥2 unique sgRNAs had average logCPM>6 and logFC>5. Gene-wise testing was performed by the Mann-Whitney test. Computer code available upon request.

For the H446 screen, extracted genomic DNA from each SVV resistant colony was used as a PCR template to amplify the lentiviral insert containing the gene-targeting sgRNA (primer # 54). The PCR product was purified then sequenced via Sanger sequencing. PCR products that contained multiple sgRNAs, as determined by Sanger sequencing, were ligated into the linearized pCR2.1 plasmid using the TA Cloning Kit (Invitrogen). The ligation reaction was transformed into DH10B cells (Invitrogen) and selected on LB agar plates supplemented 100 µg/mL carbenicillin (Fisher). Colonies from each transformation plate were isolated, amplified, and sequenced by Sanger sequencing.

**CRISPR Secondary Screens**

Individual targeting sgRNAs were cloned into the lentiCRISPRv2 plasmid as previously described(149, 152). The sgRNA plasmids were individually transfected into HEK 293T/17 cells and transduced in WT H446 or HAP1 cells as described above. Transduced H446 cells were selected with 0.5 µg/mL puromycin. Transduced HAP1 cells were selected with 1.0 µg/mL puromycin. Cells were allowed to grow for at least 7 days post transduction to allow for complete gene knockout. Cell viability was assessed by AlamarBlue fluorescent cell viability dye (ThermoFisher Scientific) as described below.
Cell Viability Assays and Analysis

Twenty-four hours prior to infection, cells (5.0×10³) were seeded into black opaque 96-well plates (Corning) in 100 µL media. Plates were infected with serial dilutions of SVV from an MOI=5,000 vp/cell to MOI= 5.0×10⁻⁵ vp/cell and incubated for 24-72 hr. Each MOI was tested in 3-6 replicate wells with uninfected cells as controls. AlamarBlue cell viability solution was added to each well and incubated at 37°C. Fluorescence emission at 590 nm was obtained after excitation at 565 nm using a Synergy Neo plate reader (BioTek) using wells containing only media as background controls. Background fluorescence values were subtracted and replicate wells averaged to determine average fluorescence and standard deviation for each MOI of SVV. The average fluorescence value at each MOI was divided by average fluorescence value of the control wells to calculate percent cell viability. Cell viability values and standard deviations were plotted against MOI of SVV using GraphPad Prism 6 software.
Results

The pooled GeCKO human sgRNA library targets over 19,000 genes within the human genome and has the ability to efficiently knock out genes using the Cas9 DNA nuclease (145, 147-149, 152). Due to the high efficiency of gene disruption in haploid cells, we performed a GeCKO screen in one of the only known human haploid cell lines, HAP1, a line derived from the chronic myelogenous leukemia (CML) cell line KBM7, which we found to be permissive to SVV at relatively high MOI (Figure 13) (139). HAP1-Cas9 cells were transduced by the pooled GeCKO library lentivirus and challenged with an SVV infection at a high MOI to select for resistant cells (Figure 14).
Figure 13. HAP1 cells are permissive to SVV. HAP1 cells were incubated with increasing MOIs of SVV for 72 h and cell viability assessed by Alamar Blue. H446 and A549 cells were assayed as positive and negative controls, respectively. Each data point represents the average of 6 replicates with error bars representing standard deviation.
Figure 14. Depiction of genome wide CRISPR knockout (GeCKO) screen workflow. After lentiviral transduction of the sgRNA library, transduced cells were selected by puromycin. Cells were then challenged with SVV to select for resistant cells.
Genomic DNA was extracted from the expanded surviving cell population and subsequently analyzed by high-throughput sequencing to determine changes in sgRNA representation compared to controls. Representation of non-targeting control sgRNAs was maintained from the plasmid pools until the end of the screen; however, notable changes were observed in targeting sgRNAs, reflecting loss of sgRNAs that target essential genes (Figure 15).
Figure 15. Plots of the relative representation of each sgRNA in the Human GeCKO v2 library at various points during the screens. 

A. Each plasmid half library was sequenced prior to lentiviral production. 

B. Representation of non-targeting sgRNAs in transduced H446 and HAP1 cells was maintained through the experiment. 

C. sgRNAs targeting both ANTXR1 (blue) and TEX2 (red) were enriched in the SVV resistant population compared to control. 

D. Cumulative distribution function plots of sgRNAs in the HAP1 screen.
demonstrates maintenance of library representation equivalent to the plasmid libraries for non-targeting sgRNAs (left panel), and loss of representation due to targeting of essential genes (right panel).
The most significantly enriched sgRNAs in the SVV selected pool were found to target the *ANTXR1* gene, which encodes the anthrax toxin receptor 1 (182). *ANTXR1* and the testis expressed 2 gene, *TEX2* were the only genes with multiple sgRNAs significantly enriched in the SVV resistant sample (Figure 16A). Highly enriched sgRNAs were individually tested in a secondary screen for the ability to confer resistance to SVV to WT HAP1 cells (Figure 16B). We observed 6 sgRNAs targeting 3 different genes that conferred SVV resistance after gene knockout in HAP1 cells, including three separate sgRNAs targeting the *ANTXR1* gene.

Both enriched sgRNAs targeting the *TEX2* gene conferred resistance as well as the sgRNA targeting the *NR2C2* gene.

To confirm our results in a cell line of immediate relevance to neuroendocrine cancers, we repeated the GeCKO screen in the highly SVV-permissive H446 SCLC cell line. GeCKO lentivirus transduced H446-Cas9 cells were challenged at MOI of 1 vp/cell for SVV. The percentage of surviving cells after SVV infection was much lower than in the HAP1 screen, allowing isolation of individual cell colonies, instead of a pooled population. Genomic DNA from each colony was extracted and individual sgRNAs were identified by Sanger sequencing (Figure 16C). In 23 of 25 resistant colonies (92%) sgRNAs targeting *ANTXR1* were present, and comprised three independent sgRNAs targeting *ANTXR1*. Each sgRNA identified in the H446 screen was tested individually in a secondary screen of WT H446 cells for the ability to confer SVV resistance (Figure 16D). All three *ANTXR1*-targeting sgRNAs identified in the screen were able to confer resistance to SVV; however, no other candidate sgRNAs altered SVV permissivity in WT H446 cells. The role of *ANTXR1* will be further examined in chapter 3.
Figure 16. Identification of ANTXR1 as an essential host determinant for SVV. 

A. The screen identified ANTXR1 (blue) and TEX2 (red) as the most significant hits in the HAP1 GeCKO screen. Non-targeting control sgRNAs are highlighted in black. 

B. HAP1 cells were transduced with individual sgRNAs identified from the HAP1 GeCKO screen. Cell viability was assayed in the absence (light grey) or presence (black) of SVV. Each bar corresponds to average of 6 replicates with error bars representing standard deviation. Dashed lines indicate WT HAP1 cell viability in the absence and presence of SVV. 

C. Table of sgRNAs identified in the H446 GeCKO screen. Twenty-five H446 colonies were isolated and the lentiviral insert sequenced by Sanger sequencing. Multiple sgRNAs were identified to target the gene ANTXR1. 

D. H446 cells were transduced with individual sgRNAs identified in the 25 H446 GeCKO screen colonies. Cell viability was tested in the absence (light grey) or presence (black) of SVV. WT H446 cell viability in the absence and presence of SVV indicated with dashed lines. Each bar corresponds to average of 6 replicates with error bars representing standard deviation.
In conclusion, recent studies have shown the exquisite power of genome-wide CRISPR screens to identify and elucidate mechanisms of cellular processes. The GeCKO sgRNA library, specifically, is capable of creating individual gene knockouts for every gene in the human genome. Multiple choices need to be considered during the design and progression of the screen to maximize the probability of a successful CRISPR screen resulting in high quality data. Using the guidelines discussed in the introductory sections of this chapter, we designed and completed two genome wide loss-of-function screens with the GeCKO sgRNA library. These screens found two genes that conferred resistance to SVV upon gene knockout. One of the genes, ANTXR1, was identified as a top hit in both screens and will be further discussed in the chapter 3.
Chapter 3. Anthrax toxin receptor 1 (ANTXR1) is the cellular receptor for SVV in neuroendocrine cancers

Picornavirus-Receptor Interactions

For every virus, a successful infection can only occur if the virus can gain access into the host cell. Although some viruses bind non-specific glycans or lipids to assist with viral entry, many viruses hijack specific cellular receptors for binding and access into the cell. This is one of the most important steps in a viral life cycle and therefore, presence of the receptor can dictate what hosts and more specifically what cells a virus can infect (183, 184). Determining the tropism of a virus is paramount to understanding its biology, which has lead to numerous important studies to identify the virus receptors for some of the most common and pathogenic viruses. For oncolytic viruses, in particular, defining the tropism of a virus can identify what cancer subtypes would be vulnerable to oncolytic therapy (39).

Some of the earliest attempts to identify viral receptors came from studies of viruses from the Picornaviridae family (185). With highly pathogenic viruses such as poliovirus (PV) and rhinovirus (common cold; HRV), it quickly became apparent after the discovery of these viruses that elucidating the method of viral entry could help to identify ways to combat viral infection. However, the limited availability of genetic, biochemical, and immunological methods to assist in identification significantly delayed discovery of viral receptors until the end of the 1900’s (185).

As of 1990, only 5 virus receptors had been identified: two of the receptors belonging to picornaviruses (185). For both PV and HRV, cell binding studies performed in multiple cell lines found that were distinct cell types that were permissive to the virus and others that were
resistant to the infection due to the inability of the virus to bind to the cells (186-190). Further characterization of cell binding conditions and enzymatic studies concluded that the receptors for both PV and HRV were glycoproteins located on the plasma membrane (187, 191-193). For PV, genetic studies using somatic cell hybrids between human (permissive) and mouse (non-permissive) cells found the receptor was located on chromosome 19 of the human genome (194). The receptor for PV, now known as poliovirus receptor (PVR) or CD155, was finally identified using human cDNA expression from ch.19 in non-permissive mouse cell lines and hybridization techniques for sequencing (195, 196). The elucidation of PVR allowed for more specific determination of PV tropism using protein expression and antibody binding studies. For HRV, binding studies on intact cells determined that rhinoviruses could be divided into two groups based on their apparent tropism (189, 197). The major group of rhinoviruses, which included HRV-14, were further found to use the same receptor as the coxsackievirus strains A13, A18, and A21 (198, 199). Using monoclonal antibody affinity chromatography and purification, the receptor for the HRV major group was initially isolated. Amino acid sequencing identified the glycoprotein to be the intercellular adhesion molecule-1 (ICAM-1). Studies confirmed ICAM-1 was also the receptor for coxsackieviruses A13, A18, and A21 (200-204).

The development of more powerful biochemical, immunological, and genetic methods as well as genome wide screens have allowed for the discovery of many receptors across multiple families of viruses. More recent studies have identified the receptors for other picornaviruses such as coxsackievirus B, enterovirus 71, and the minor group of HRV, among others (205-208). The majority of known picornavirus receptors are single transmembrane proteins on the plasma membrane. One common theme emerging among
known picornavirus receptors is the preference in utilizing proteins from the immunoglobulin superfamily (IgSF) as receptors. PVR, ICAM-1, and the coxsackievirus-adenovirus receptor (CAR) for some coxsackievirus type B, all contain IgSF domains that interact directly with the corresponding virus (209). However, based on currently identified picornavirus receptors, the use of IgSF proteins is not universally conserved.

Advances in crystallography and cryo-electron microscopy (cryo-EM) have allowed for researchers to visualize not only the structure of a virus but also the direct interaction between viruses and their receptors. To date, structural studies have been performed with PV, major group and minor group HRVs, and coxsackievirus type A and B among others (210-214). Based on structural studies, a consensus receptor attachment hypothesis, known as the canyon hypothesis, was proposed for picornaviruses (69, 215). Many picornaviruses were determined to have a deep depression, or canyon, in their capsid surface surrounding the five-fold vertex. Because these canyons are inaccessible to neutralizing antibodies, researchers hypothesized that the canyon allowed for viruses to conserve these residues for receptor recognition and attachment (69, 215). Binding studies have further suggested that receptors, IgSF protein receptors in particular, that bind within the canyon also lead to conformational changes that destabilize the viral capsid and ultimately lead to viral uncoating (216-218). However, there are receptors, almost all non-IgSF proteins, that bind certain picornaviruses outside of the canyon (209, 219). Furthermore, the resolution of cardiovirus structures and others have found the canyon depression on the viral capsid is not universally conserved. Instead, cardioviruses have a “sharp pit” where the center of the canyon is located on other picornaviruses. There is some evidence to support receptor binding within the pit; however, more research is needed to confirm across other picornaviruses (220, 221). In both instances,
binding to the receptor ultimately leads to viral entry or viral uncoating. Viral entry most likely occurs by receptor-mediated endocytosis, but other pathways may be used (222).

The structure of the SVV-001 viral capsid was resolved shortly after the initial discovery of the virus. Similarly to closely related cardioviruses, SVV-001 was determined to have a “sharp pit” depression at the five-fold vertex instead of a deep canyon. Other structural differences in the viral capsid helped to determine that SVV-001 should be classified into a new genus (101). Until now, the receptor for SVV-001 was unknown and therefore, studies have not been performed to characterize the interaction between SVV-001 and its receptor.
Study Introduction

In the previous chapter, we described two genome-wide loss of function screens with the genome-scale CRISPR/Cas9 knock-out (GeCKO) human single-guide RNA (sgRNA) library to identify essential host factors for SVV infection in two permissive cell lines. The screens and studies described here have identified anthrax toxin receptor 1 (ANTXR1) as an essential gene for SVV infection and the high affinity cellular receptor for SVV. We demonstrate that SVV interacts directly and specifically with ANTXR1, that this interaction is required for SVV binding to permissive cells, and that ANTXR1 expression is essential for infection. Lastly, we identified the region of the SVV capsid responsible for receptor recognition with cryo-electron microscopy (cryo-EM) of the SVV-ANTXR1-Fc complex. These studies define a clinically tractable predictive biomarker of SVV permissivity, and identify ANTXR1 as the high-affinity cellular receptor for SVV in neuroendocrine cancers.
Materials and Methods

Reagents and Bacterial Strains

Polymerase chain reactions (PCRs) were carried out using a GeneAmp PCR System 9700 Thermocycler (Applied Biosystems). PCR fragments were purified using QIAquick PCR Purification Kit (Qiagen). Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen). Competent DH10B and Stbl3 cells were purchased from Invitrogen. Plasmids were isolated and purified from bacteria using QIAquick Spin Miniprep Kit (Qiagen). Sanger sequencing for individual clones and plasmids was performed by Genewiz, Inc. Nucleotide and protein sequence alignments were performed in Geneious Pro 4.7.6.

Cell Lines and Viruses

All cell culture media was produced by the Memorial Sloan Kettering Cancer Center (MSKCC) Media Prep core facility. HAP1 cells were purchased from Haplogen GmbH and maintained in Iscove’s Modified Dulbecco’s Medium (IMDM) supplemented with 10% fetal calf serum. TC-71 cells were obtained from Children’s Oncology Group (COG) Cell Culture Repository and maintained in IMDM supplemented with 10% fetal bovine serum and 1× ITS supplement. All other cell lines used in this study were purchased from ATCC. HEK 293T/17 cells were maintained in high glucose Dulbecco’s Modified Eagle Medium (DMEM HG) supplemented with 10% fetal calf serum and 1 mM sodium pyruvate. All additional lines were maintained in RPMI 1640 supplemented with 10% fetal calf serum and 10 mM HEPES. All cell lines are routinely confirmed by STR analysis and confirmed mycoplasma
negative by DDC Medical. SVV and SVV-GFP was cultured, purified, and titered as previously described (59, 62).

**Individual sgRNA Plasmids**

The lentiCRISPR v2 plasmid, which expresses a single sgRNA under the hU6 promoter and the WT Cas9 nuclease under the EFS promoter, was a gift from Feng Zhang (Addgene plasmid #52961). Oligos containing the gene targeting sgRNA with 5’ overhang BsmBI digestion sites were synthesized by Sigma Aldrich (primers #22-53, Table 3, pg. 89). The oligos were annealed and inserted into the lentiCRISPR v2 backbone as described previously (149, 152). sgRNA sequences and plasmids were confirmed by Sanger sequencing.

**Inducible ANTXR1 Expression Plasmid Construction**

To create an inducible ANTXR1 expression lentiviral plasmid, we obtained a plasmid expressing a full-length ANTXR1 cDNA as a generous gift from Drs. Stephen Leppla and Shihui Liu from NIH/NIAID. The plasmid (ANTXR1-HA), which constitutively expresses the ANTXR1 cDNA fused to a C-terminal influenza virus hemagglutinin (HA) tag, was used as a template in PCR with primers incorporating Gateway attB1 and attB2 cloning sequences (primer # 57-58). The PCR product was purified and subsequently used in a BP reaction with the Gateway destination vector, pDONR221 (Invitrogen), transformed into DH10B cells, and selected on LB agar plates supplemented with 50 µg/ml kanamycin (Sigma Aldrich). The plasmid containing the PCR fragment was purified and used in an LR reaction with pInducer20, a gift from Stephen Elledge (Addgene plasmid #44012). The recombinant
plasmid was transformed into Stbl3 cells and selected on LB agar plates supplemented with 100 µg/ml carbenicillin (Sigma). The purified plasmid, containing a doxycycline (Dox) inducible ANTXRI cDNA with the HA fusion tag, was confirmed by Sanger sequencing and used to produce lentivirus for the creation of stable cell lines.

**Lentivirus Production**

The lentiviral packaging plasmids pMD2.G (Addgene plasmid #12259) and psPAX2 (Addgene plasmid #12260) were gifts from Didier Trono. All transfections of lentiviral plasmids were performed as follows unless otherwise stated: Lentiviral plasmids were transfected at a 3:2:1 DNA ratio of lentiviral plasmid:psPAX2:pMD2.G in 1 mg/mL polyethylinimine (PEI; Sigma Aldrich) at a 2:1 PEI:DNA ratio in OptiMEM (Gibco). Media was changed 16 hr post transfection. Seventy-two hours post transfection the virus containing media was harvested and filtered through a 0.45 µm PDVF syringe filter (Millipore) to remove cell debris. Lentivirus aliquots were stored at -80°C.

**Lentiviral Transductions**

All lentiviral transductions were performed as follows unless otherwise stated: Cells (1.0×10^6) to be transduced were plated in a 75 cm² flask the day before transduction. Lentivirus was thawed on ice and added to OptiMEM supplemented with 32 µg/mL polybrene. After the virus-OptiMEM mix was added to the cells, additional media was added to bring the final polybrene concentration to 8 µg/mL. Media was changed 24 hr after transduction to remove polybrene. Media supplemented with 0.5 µg/mL puromycin (Sigma Aldrich) or 6 µg/mL blasticidin (Fisher Scientific) was changed 48 hr after transduction to
select lentiCRISPRv2 or lentiCas9-Blast transduced cells, respectively. Transduced cells were maintained in media containing either puromycin or blasticidin. For Dox inducible ANTXR1 lentivirus (pInducer20-ANTXR1), the SCLC H69 and H146 cell lines were transduced and maintained in tetracycline-free (tet-free) media supplemented with 500 µg/mL G-418 (Thermo Fisher).

Cell Viability Assays and Analysis

Twenty-four hours prior to infection, cells (5.0×10^3) were seeded into black opaque 96-well plates (Corning) in 100 µL media. Plates were infected with serial dilutions of SVV from an MOI=5,000 vp/cell to MOI= 5.0×10^{-5} vp/cell and incubated for 24-72 hr. Each MOI was tested in 3-6 replicate wells with uninfected cells as controls. AlamarBlue cell viability solution was added to each well and incubated at 37°C. Fluorescence emission at 590 nm was obtained after excitation at 565 nm using a Synergy Neo plate reader (BioTek) using wells containing only media as background controls. Background fluorescence values were subtracted and replicate wells averaged to determine average fluorescence and standard deviation for each MOI of SVV. The average fluorescence value at each MOI was divided by average fluorescence value of the control wells to calculate percent cell viability. Cell viability values and standard deviations were plotted against MOI of SVV using GraphPad Prism 6 software.

Identification of ANTXR1 indels in ANTXR1 KO lines

Extracted genomic DNA from WT H446 cells and ANTXR1 KO mutant clones was used as a PCR template to amplify the target of the ANTXR1 sgRNA, exon 2 of the ANTXR1 gene,
using sequence specific primers (primer #55-56). The PCR product was ligated into the linearized pCR2.1 plasmid using the TA Cloning Kit (Invitrogen). The ligation reaction was transformed into DH10B cells (Invitrogen) and selected on LB agar plates supplemented 100 µg/mL carbenicillin (Fisher Scientific). Colonies from each transformation plate were isolated, amplified, and sequenced by Sanger sequencing. Exon 2 sequences from ANTXR1 KO lines were compared to WT H446 exon 2 sequences and ANTXR1 gene reference sequence (NG_012649.1; Pubmed) to identify indels in each cell line.

**ANTXR1-KO Lines**

All cell lines were transduced with ANTXR1.3 sgRNA lentiCRISPRv2 lentivirus as described above. Transduced cells in each cell line were selected with 1.0 µg/mL puromycin. Cells were allowed to grow for at least 7 days post transduction to allow for complete gene knockdown. As a negative control cell line, WT H446 cells were similarly transduced with lentiCRISPRv2 lentivirus containing an EGFP targeting sgRNA, a gift from Feng Zhang (Addgene plasmid # 51764), and selected with puromycin.

**ANTXR1 Expression Experiments**

Unless otherwise stated, cells were transfected and analyzed as follows: ANTXR1 rescue experiments were performed with the ANTXR1-HA expression plasmid. Cells were plated in tissue culture treated 6-well plates 24 hr prior to transfection. Cells were transiently co-transfected with ANTXR1-HA and pLenti6 W118-mCherry, which constitutively expresses the fluorescent protein mCherry, in 1 mg/mL PEI in OptiMEM with untransfected cells as controls. Media was changed 16 hr post transfection. The cells were then harvested for
Western blot analysis or challenged with SVV-GFP for flow cytometry analysis as described below. For Western blot lysates, transfected cells were harvested 48 hr post transfection and pelleted by centrifugation. Cell pellets were lysed in radioimmunoprecipitation assay (RIPA) buffer (Pierce) supplemented with 1X Halt Protease and Phosphatase Inhibitor Cocktail (Pierce) and subsequently clarified by centrifugation. Protein lysates were quantified using the BCA protein assay kit (Pierce) and prepared for Western blot analysis by boiling in for 10 min at 90 °C in NuPAGE sample reducing agent and LDS sample buffer (Invitrogen). Western blots analysis was performed as described below.

For re-expression experiments in SCLC ANTXR1 KO lines, cells were transiently co-transfected with ANTXR1-HA and pLenti6 W118-mCherry expression plasmids at a 10:1 molar ratio in 1 mg/mL PEI in OptiMEM with cells transfected with mCherry alone as controls. Cells were challenged with SVV-GFP at the TCID\textsubscript{50} for each cell line 48 hr post transfection and harvested for analysis 6 hr post SVV-GFP infection. For expression experiments in non-permissive SCLC cell lines, pInducer20-ANTXR1 transduced H69 or H146 cells were seeded in 6-well plates 16 hr prior to the start of the experiment. Cells were maintained in tet-free media alone or supplemented with 1 µg/mL doxycycline for 72 hr prior to the addition of SVV-GFP. Cells were incubated with SVV-GFP for 6 h then harvested for analysis.

\textit{SVV-GFP Infections}

Unless otherwise stated, cells were seeded in a tissue culture treated well plate (Corning) 24 hr prior to infection. Plates were infected with SVV-GFP at an MOI=5.0 vp/cell and incubated at 37 °C for 8 or 16 hr. NucBlue Live ReadyProbe reagent (Invitrogen) was added
to each well and incubated at 37 °C for 20 min. Images of cells were obtained using an EVOS FL Auto fluorescence microscope (Invitrogen). For IFN response activity assays, DMS79 and H1618 cells were treated with media alone, media supplemented with 4 μM of histone deacteylase (HDAC) inhibitor MS-275 (Selleck) or Vorinostat (SAHA; LC labs), 25 Units/mL IFN-α (Thermo) or IFN-β (R&D Systems), or 25 U/mL IFN-α/β and corresponding 5 μg/mL IFN-α (Thermo) or IFN-β antibody (R&D Systems) for 24 hr at 37 °C. H446, H82, and H1618 cells were treated with media alone or media supplemented with 25 U/mL IFN-β for 24 hr at 37 °C. Cells were then challenged with SVV-GFP (H1618: MOI= 0.5 vp/cell; DMS79, H446, H82: MOI=0.1 vp/cell) for 16 hr at 37 °C and subsequently analyzed by flow cytometry. For blocking experiments, SVV-GFP (MOI=5) was incubated with 5 μg/mL ANTXR1-Fc or ANTXR2-Fc chimera, IgG-Fc (Sino Biological), or control (R&D Systems) on ice for 1 h and subsequently added to cells for 16 h at 37 °C. For blocking experiment with PNGase F, 5 μg/mL ANTXR1-Fc, IgG-Fc, or control buffer was pre-incubated with PNGase F (1000 U; NEB) at 37 °C for 3 hr then subsequently incubated with SVV-GFP (MOI=5) on ice for 1 h. Each condition was then added to cells for 16 h at 37 °C.

**SVV-Cy5 Binding Experiment**

SVV was incubated with the amine-reactive Cy5 dye (GE Healthcare) in sodium carbonate buffer (pH 9.3) for 1 hr at room temperature (RT). Excess dye was removed by filtration through gel filtration columns (GE Healthcare) in HEPES buffer. Virus aliquots were stored at -80°C. WT, ANTXRI KO, and TEX2 KO H446 cells were incubated with SVV-Cy5 for 30 min at 37°C in OptiMEM. The non-permissive SCLC cell line, DMS114, was used as a
negative control. Cells were then washed with FACS buffer three times before staining with LIVE/DEAD Fixable Aqua Dead Cell Stain Kit and fixed as described above. Samples were run on a BD LSR II Flow Cytometer (Becton Dickinson) using unstained cells and cells incubated with OptiMEM alone as controls. All compensation and gating were performed with FlowJo analysis software (TreeStar) as described below.

**Co-Immunoprecipitations (Co-IP)**

Magnetic Protein G Dynabeads (Invitrogen) were used for all immunoprecipitation experiments. Unless otherwise stated, Dynabeads and Dynabead-protein complexes were washed three times with PBS, pH 7.4 supplemented with 0.02% Tween-20 (Sigma Aldrich). Dynabeads were immobilized for manipulation and washing using a DynaMag magnet (Life Technologies). Proteins were eluted by boiling Dynabead-protein complexes for 10 min at 90 °C using RIPA buffer supplemented with NuPAGE sample reducing agent and LDS sample buffer. For initial ANTXR1-Fc and ANTXR2-Fc co-IP experiments, serial dilutions of Fc chimera proteins (0.25 µg) in PBS, pH 7.4 were incubated with 1 µL of 30 mg/mL Dynabeads for 10 min at room temperature. Dynabead-Fc complexes were washed and subsequently incubated with SVV (2.0×10^{10} vp) for 2 h at 4 °C. Triplicate washes were repeated and dynabead-protein complexes were then subjected to protein elution. For co-IP using high stringency washes, PBS, pH 7.4 supplemented with 0.02% Tween-20 and increasing amounts of NaCl from 125 mM to 2 M were used to wash the Dynabead-protein complexes after SVV addition. For PNGase F studies, ANTXR1-Fc (0.5 µg) in PBS, pH 7.4 were incubated with 2 µL of 30 mg/mL Dynabeads for 10 min at room temperature. Dynabead-Fc complexes were washed and subsequently incubated with PNGase F (1000 U)
for, 0, 4, 8, 16, or 24 hr at 37 °C. Dynabead-Fc complexes were washed and incubated with SVV-001 (4.0×10^{10} vp) and processed as described above. PNGase F samples were eluted in double the volume as other samples and split into equal samples for parallel Western blot and silver stain analysis.

**Western Blotting**

Eluted Dynabead proteins or protein extracts were resolved on a 4-12% Bis-Tris polyacrylamide gel with MOPS running buffer (Life Technologies) and transferred to a polyvinylidene difluoride (PVDF) membrane (Millipore). For co-IP experiments, membranes were blotted with rabbit antisera against SVV (Neotropix) For *ANTXR1* transfection cell lysates, membranes were blotted with commercial primary antibodies against the HA tag (Cell Signaling cat. No. 3724S) or glyceraldehyde phosphate dehydrogenase as loading control (GAPDH; Santa Cruz cat. no. sc-25778). Immunoblotting was performed using horseradish peroxidase–conjugated secondary antibodies (Cell Signaling) and detection by chemiluminescence (GE Life Sciences).

**Silver Stain**

Eluted Dynabead proteins or protein extracts were resolved on a 4-12% Bis-Tris polyacrylamide gel with MOPS running buffer (Life Technologies). Gels were then silver stained using the Silver Stain Kit for Mass Spectrometry (Pierce). Images of silver stained gels were taken using a E-Gel Imager (Life Technologies).

**Flow Cytometry Analysis**
WT and *ANTXR1* KO cell lines were seeded in tissue culture treated 6-well plates 24 h prior to SVV-GFP infection. Cells were infected with SVV-GFP at the TCID$_{50}$ for each cell line and incubated at 37 °C for 6-16 hr with uninfected cells as controls. Cells were subsequently harvested using Accutase enzyme cell detachment media (Gibco), pelleted by centrifugation, and washed with sterile PBS, pH 7.4 supplemented with 2% FCS and 0.5 mM EDTA (FACS buffer). Cells were stained with LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Invitrogen) for 30 min at 4 °C, washed with FACS buffer, then fixed in 4% paraformaldehyde solution for 10 min at 4 °C. After a final wash with FACS buffer, cells incubated with SVV-GFP were run on a BD LSR II Flow Cytometer (Becton Dickinson) using unstained cells and cells incubated with media alone as controls. All experimental samples were collected and performed in triplicate. Additional gating and analysis was performed with FlowJo analysis software (TreeStar). Analyzed data and standard deviations were plotted using GraphPad Prism 6 software. Unpaired two-sided t tests were performed where applicable to determine statistical significance. Asterisks represent significance levels as follows: * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001, and **** p ≤ 0.0001.

**Gene expression analysis**

Normalized gene expression data for 1,037 cancer cell lines was downloaded from the Cancer Cell Line Encyclopedia (http://www.broadinstitute.org/ccle/home). To determine the appropriate cutoff for cell lines expressing *ANTXRI*, local modes in the density distribution of *ANTXRI* expression were identified and designated the lowest mode as non-expressed. The standard deviation of this peak was then determined and an expression cutoff equal to 10 standard deviations above the mode was set, based on the work of Zilliox et al. Gene
expression analysis was performed using the R statistical programming environment and the Bioconductor suite of tools. Differentially expressed genes were identified using LIMMA to fit a linear model to each gene and generate moderated t-statistics using an empirical Bayes approach. Gene set enrichment analysis was performed using CAMERA, a purely competitive gene set testing approach(176). Sample-wise enrichment was determined using GSVA(223).

Cryo-Electron Microscopy

Equal volumes of virions at 0.2 mg/ml and ANTXR1 at 1 mg/ml were mixed, giving a ratio of ~10:1 receptors per binding site. The samples were mixed and kept for 90 min at 37°C and transferred on ice for another 90 min. Specimens were prepared by applying 3 µL of purified virus on glow discharged Quantifoil holey carbon grids (Quantifoil Micro Tools GmbH, Grossloebichau/Jena, Germany). The excess buffer was blotted and the grid was flash plunged into liquid ethane using a Leica KF80 cryo fixation device (C. Reichert Optische Werke AG, Vienna, Austria). Grids were loaded onto a Gatan 914 Cryoholder (Pleasanton, CA, USA). Images were collected on a JEOL JEM2200FS microscope (JEOL Ltd, Tokyo, Japan) operated at 200 kV using minimal dose conditions with an electron dose of ~30 electrons/Å². An in-column omega energy filter was used to improve image contrast by zero-loss filtering with a slit width of 25 eV. Automated data collection was carried out using SerialEM software. The micrographs were recorded at a defocus between 1 and 3 µm, on a 4×4 k CMOS camera (TVIPS; Gauting, Germany) at a calibrated magnification of 50,000 corresponding to a pixel size of 3.12 Å.
A number of 400 individual virus particles were selected from micrographs using the E2BOXER software(224). Contrast Transfer Function parameters were calculated using CTFFIND3(225), and micrographs with poor CTF estimates were discarded. Orientation, classification and refinement were done in Relion(226) using as initial reference a strongly low pass version of the SVV atomic model(101). By calculating the Fourier shell correlation between two halves of the data set, the resolution of the map was estimated to be 18 Å. The reconstructed map was visualized using Chimera(227).
Results

We assessed the genomic sequence of ANTXRI in clones isolated from the H446 screen and found that all 5 ANTXRI KO clones contained insertions or deletions (indels) in exon 2 of the ANTXRI gene. These indels would lead to a frame shift mutation and premature stop codon, confirming that these changes should lead to complete loss of ANTXRI protein (Figure 17A). We also confirmed the loss of SVV permissivity of the ANTXRI KO cells using a cell viability assay with WT H446 cells and non-permissive A549 cells as positive and negative controls, respectively (Figure 17B). After a 72 h incubation period with SVV, we observed a significant loss of viability with increasing MOI of SVV in WT H446 cells. Indicative of highly SVV resistant cells, all ANTXRI KO lines as well as A549 cells showed no significant change in cell viability with MOI of SVV over 5 logs higher than effective exposures for WT H446 cells.

To determine whether ANTXRI is essential for SVV infection in additional neuroendocrine cancer cell lines, we generated ANTXRI KO lines in the SCLC cell lines H446, LX22cl, and H82, as well as HAP1 and the SVV-permissive pediatric cancer cell lines Y79 and TC-71. Each ANTXRI KO line was challenged with an infectious SVV reporter virus that expresses GFP within the viral polyprotein (SVV-GFP)(20, 62). Cells were analyzed by flow cytometry using the corresponding WT cell line as a positive control (Figure 17C). As a negative control, we created an H446 cell line stably expressing an sgRNA targeting EGFP, to confirm loss of permissivity was due to the targeting ANTXRI sgRNAs and not off-target effects (Figure 17D). In all cases ANTXRI KO profoundly decreased SVV-GFP infection
≥70% in the KO cell lines compared to the corresponding WT lines. \textit{ANTXR1} gene knockout leads to a loss of SVV permissivity in permissive cell lines of multiple tumor types.
Figure 17. Knockout of the \textit{ANTXR1} gene leads to the loss of SVV permissivity. A. Summary table of \textit{ANTXR1} indels in five selected H446 \textit{ANTXR1} KO colonies from the GeCKO screen. B. Three of the H446 \textit{ANTXR1} KO cell lines were challenged with increasing MOIs of SVV for 72 h. Cell viability was determined via alamarBlue. WT H446 cells and the non-permissive NSCLC cell line A549 were used as positive and negative controls, respectively. Each data point represents the average of 6 replicates with error bars representing standard deviation. C. WT permissive SCLC and pediatric cell lines were transduced with a sgRNA targeting \textit{ANTXR1}. WT (light grey) and \textit{ANTXR1} KO (black) cells were challenged with SVV-GFP and analyzed by flow cytometry. D. H446 WT (light grey) and \textit{GFP} sgRNA containing H446 (black) cells were challenged with SVV-GFP and analyzed by flow cytometry. Each bar in c) and d) represents the average of 3 replicates with error bars representing standard deviation.
We next sought to determine whether *ANTXR1* expression levels in cell lines is predictive of permissivity using publicly available gene expression data of the 1,037 cell lines in the Cancer Cell Line Encyclopedia (CCLE)(228). We determined an expression cutoff based on the distribution of cell lines in CCLE(229). Approximately 37% of cell lines fell below the expression cutoff (Figure 18A). Of the cell lines in the CCLE, 81 have been previously assessed for permissivity. Of these lines, biased toward inclusion of neuroendocrine cancer lines, 20 were found to be permissive. *ANTXR1* expression was significantly associated with permissivity (*p*=0.0023, Fisher’s exact test). Most strikingly, none of the 20 permissive cell lines lacked expression of *ANTXR1*, supporting the hypothesis that *ANTXR1* is a required host factor for SVV infection.

While *ANTXR1* expression appears to be a requirement for SVV permissivity, the CCLE dataset suggests that it is not sufficient: 42/62 (67.7%) of *ANTXR1*-expressing cell lines analyzed for permissivity were reported to be non-permissive. We next sought to identify meaningful gene expression changes between *ANTXR1*-expressing permissive and non-permissive classes. We used competitive gene set enrichment to identify significantly differentially expressed gene sets from the Reactome database(230, 231). We identified 7 gene sets, all of which were significantly down-regulated in permissive cell lines expressing *ANTXR1* (Figure 18B). The most significant gene set was INTERFERON_ALPHA_BETA_SIGNALING, in which 34/44 (77%) of genes were significantly down-regulated in permissive cell lines. The enrichment for this gene set (*q*=0.0046) can be visualized in Figure 18C. A sample-wise analysis of gene expression was performed to see whether the gene set enrichment we observed was driven by cell lines derived from a particular tumor histology and we found that lack of expression of these gene
sets was enriched among SCLC and neuroblastoma cell lines (Figure 18D) (223). The sample-wise analysis of gene expression is shown in Figure 19A.
Figure 18. ANTXR1 expression is significantly associated with permissivity. A. Scaled log2 ANTXR1 gene expression of permissive (top), non-permissive (middle) and all cell lines in the CCLE (bottom). ANTXR1 expression was significantly associated with permissivity ($p = 0.0023$; Fisher’s exact test). B. Competitive gene set testing analysis revealed 7 highly significantly differentially expressed gene sets, all down-regulated in permissive cell lines. C. An enrichment barcode plot depicting the negative enrichment of type I interferon signaling genes in permissive cell lines ($q=0.0046$). D. Sample-wise enrichment scores were calculated.
for the top enriched gene set and plotted based on the histology of the tumor of origin. SCLC and neuroblastoma stand out as lacking genes involved in type I interferon signaling.
To experimentally confirm the gene expression data, we tested the ability to confer changes in SVV permissivity by altering the IFN pathway activity in two SCLC cell lines, H1618 and DMS79, determined to be non-permissive by cell viability assays. A previous study found that pre-treatment of cells with histone deacetylase (HDAC) inhibitors conferred increased sensitivity to vesicular stomatitis virus due to decreased cellular IFN responses (232). We exposed cells to HDAC inhibitors, MS-275 or Vorinstat (SAHA), to decrease the IFN pathway and IFN-α or IFN-β to activate the IFN pathway for 24 hr prior to challenge with SVV-GFP for 24 hr. The cells were then analyzed by flow cytometry (Figure 19 B and C). Surprisingly, both cell lines contained a small SVV permissive subpopulation at baseline although they had previously been determined to be non-permissive. In both cell lines, there was a significant increase in GFP positive (GFP⁺), or SVV permissive, cells after pre-treatment with either HDAC inhibitor, in correlation with data from Nguyen, T.L., et al. (232). Conversely, after treatment with IFN-α (H1618) or IFN-β (DMS79), there was a decrease in GFP⁺ cells in both cell lines. The change in permissivity could be partially rescued by adding an IFN-α or IFN-β antibody in combination with IFN-α or IFN-β pre-treatment, respectively, confirming the change in permissivity was due to the presence of the IFNs prior to SVV-GFP exposure. These studies established that SVV permissivity in cell lines with low SVV permissivity can be altered by the activity of the IFN pathway within the cells. We next sought to determine if permissivity could be affected by IFN pathway activation in highly permissive cells with significantly downregulated IFN pathway genes. We exposed H446 and H82 cells to IFN-β for 24 hr prior to treatment with SVV-GFP for 24 hr and subsequently analyzed the cells by flow cytometry (Figure 19D). In contrast to control H1618 cells that showed a significant decrease in GFP⁺ cells after IFN-β treatment, the level
of SVV permissive H446 and H82 cells was not severely altered by IFN-β treatment. This observation confirms that permissivity levels can be altered by activating or inhibiting the IFN response pathways, but only in cells with intact or active IFN pathways, as H446 and H82 cells show a drastic downregulation in the IFN pathway genes. Taken together, these results suggest that robust permissivity to SVV requires both expression of the cellular receptor ANTXR1 and down-regulation of expression of antiviral INF signaling genes.
Figure 19. Activity of IFN response pathways can alter SVV permissivity in low SVV permissive cell lines. A. Sample-wise analysis of gene expression across ATXRI\(^+\) cell lines. Permissivity of each cell line was previously determined by cell viability assays. B-C. SCLC cells determined to be non-permissive by cell viability assays, H1618 (B) and DMS79 (C) cells (uninfected: black; SVV-GFP only: green) were pre-treated with HDAC inhibitors, MS-275 (dark blue) or Vorinstat (SAHA; light blue), IFN-\(\alpha\) (H1618)/ IFN-\(\beta\) (DMS79) (red), or IFN-\(\alpha/\beta\) with a corresponding IFN-\(\alpha/\beta\) antibody (light pink) prior to incubation with SVV-GFP. Cells were analyzed by flow cytometry. D. H446, H82, and H1618 cells were pre-treated with IFN-\(\beta\) (light blue) prior to exposure to SVV-GFP using untreated cells as controls (dark blue). Cells were analyzed by flow cytometry. Each bar in B-D represents the average of 3 replicates with error bars representing standard deviation.
To confirm the specificity of the *ANTXR1* sgRNAs, we evaluated whether exogenous re-expression of *ANTXR1* could rescue permissivity to SVV in *ANTXR1* KO cells. We co-transfected three H446 *ANTXR1* KO lines with an ANTXR1-HA expression plasmid and an mCherry fluorescent protein expression plasmid and challenged the cells with SVV-GFP 16 h post transfection (Figure 20A). Compared to untransfected *ANTXR1* KO cells that did not show any GFP positive (GFP\(^+\)) cells, *ANTXR1* KO cells transfected with the ANTXR1-HA expression plasmid consistently showed GFP\(^+\) cells, indicative of a productive SVV-GFP infection and rescue of SVV permissivity. To further test the importance of ANTXR1 expression in permissive cells, we co-transfected the H446 and LX22cl *ANTXR1* KO lines with the ANTXR1-HA and mCherry expression plasmids, and subsequently incubated the cells with SVV-GFP. Cells were analyzed by flow cytometry and gated to select for transfected cells (mCherry\(^+\)). Compared to WT mCherry\(^+\)/GFP\(^+\) cells, we observed a significant decrease in the mCherry\(^+\)/GFP\(^+\) population in *ANTXR1* KO H446 and LX22cl cells to 2.58 ± 0.91% and 10.50 ± 3.26%, respectively (Figure 20B). Upon transfection with the ANTXR1-HA expression plasmid, mCherry\(^+\)/GFP\(^+\) populations were rescued to 110.4 ± 9.9% and 35.5 ± 7.8% in H446 and LX22cl *ANTXR1* KO cells, respectively, relative to WT cells. Expression of the ANTXR1-HA fusion protein was confirmed in each ANTXR1-HA transfected cell line by immunoblot using an HA tag specific antibody. Re-expression of ANTXR1 protein in *ANTXR1* KO cell lines is sufficient to rescue SVV permissivity.
Figure 20. Re-expression of ANTXR1 reconstitutes SVV permissivity. Cells were co-transfected with the ANTXR1-HA and mCherry expression plasmids (A,B,C) A. Three H446 ANTXRI KO cell lines were transfected then challenged with SVV-GFP. Scale bar represents 100 µm. B. H446 and LX22cl ANTXRI KO cell lines were transfected, challenged with SVV-GFP, and analyzed by flow cytometry. mCherry transfected WT and ANTXRI KO cells were used as positive and negative controls, respectively. Each bar represents the average of 3 replicates with error bars representing standard deviation. C. H69 and H146 cells were transduced with a Dox-inducible ANTXR1-HA expression lentivirus. WT and ANTXRI expressing cells were incubated in absence or presence of 1 µg/mL doxycycline for 72 h, challenged with SVV-GFP, and analyzed by flow cytometry. Each bar represents the average of 3 replicates with error bars representing standard deviation.
We next sought to determine whether expression of ANTXR1 protein was sufficient to increase the permissivity of the non-permissive SCLC cell lines H69 and H146, which do not express the gene. After transduction with a doxycycline inducible ANTXR1-HA expression lentivirus, we incubated WT and ANTXR1 expressing H69 and H146 cells in the presence or absence of 1 µg/mL doxycycline for 72 h, challenged with SVV-GFP, and analyzed by flow cytometry (Figure 20C). Parental H69 and H146 cells showed GFP+ populations under 0.25% regardless of doxycycline, as expected. ANTXR1-transduced H69 and H146 cells showed similar GFP+ populations of 1.17 ± 0.33% and 0.22 ± 0.07%, respectively, in the absence of doxycycline. Upon doxycycline treatment, both ANTXR1-transduced H69 and H146 cells showed a significant increase in SVV-GFP infected cells to 7.46 ± 0.17% and 18.3 ± 0.20%, respectively. Expression of the ANTXR1-HA protein in the doxycycline induced cells was confirmed by Western blot. These data confirm that the ANTXR1 protein is sufficient to induce permissivity in SVV-resistant SCLC cell lines.

As ANTXR1 is a transmembrane protein and required for SVV infection in various permissive SCLC cell lines, we sought to determine whether ANTXR1 interacts directly with SVV. We utilized an ANTXR1-Fc chimera or a control isotype IgG1 Fc protein for co-immunoprecipitation studies. After incubating Fc-bead complexes with SVV, all bound proteins were eluted and analyzed by Western blot using SVV rabbit antisera (Figure 21A). In all serially diluted ANTXR1-Fc samples incubated with SVV, we observed viral protein bands as well as a decrease in intensity of the bands corresponding to a decrease of bound ANTXR1-Fc protein. We did not detect any SVV protein bands in samples incubated with the IgG1 Fc isotype control or samples not incubated with SVV. After confirming a direct interaction, we repeated the ANTXR1-Fc chimera co-immunoprecipitation studies in the
presence of increasing amounts of sodium chloride (NaCl) to investigate the strength of the interaction \textit{in vitro} under high ionic strength (Figure 21B). The intensity of viral protein bands did not change significantly with increasing salt concentration up to 2M NaCl. As ANTXR1 has high sequence similarity to the high-affinity anthrax receptor, ANTXR2, we performed the co-immunoprecipitation with the ANTXR2-Fc chimeric protein using the ANTXR1-Fc protein as a positive control(182). We did not observe any bands corresponding to viral proteins in ANTXR2-Fc samples incubated with SVV, indicating the absence of an interaction between the extracellular domain of ANTXR2 and SVV (Figure 21C). Bands corresponding to viral protein were observed only in ANTXR1-Fc samples. Our results indicate that ANTXR1, and not ANTXR2, can directly interact with SVV in a high affinity and stable interaction.
Figure 21. The direct SVV-ANTXR1 interaction is specific and of high affinity in co-immunoprecipitation assays. 

A. SVV was co-immunoprecipitated with decreasing amounts of an ANTXR1-Fc chimera. Bound proteins were eluted and analyzed by Western blot using an anti-SVV antibody. Input SVV was immunoblotted as a positive control.

B. SVV was co-immunoprecipitated with the ANTXR1-Fc chimera. Washes were performed with increasing concentrations of NaCl up to 2 M. Bound proteins were eluted and analyzed as described in A.

C. SVV was co-immunoprecipitated with the ANTXR1-Fc chimera or decreasing amounts of ANTXR2-Fc chimera and analyzed as described in A.
We next sought to determine if the interaction between the SVV and ANTXR1-Fc or ANTXR2-Fc chimera could attenuate a cellular SVV infection. We incubated SVV-GFP with the ANTXR1-Fc, ANTXR2-Fc, or IgG1-Fc protein prior to an overnight incubation with WT H446 cells and subsequent analysis by fluorescence microscopy (Figure 22A). Cells incubated with SVV-GFP and IgG1-Fc or ANTXR2-Fc protein showed high levels of GFP+ cells indicative of a productive SVV infection. Cells incubated with SVV-GFP and ANTXR1-Fc protein showed no detectable GFP+ cells, indicating a substantial lack of SVV-GFP infection in these cells. These results demonstrate that only exogenous ANTXR1 protein and not ANTXR2 protein is able to block a cellular SVV-GFP infection, and further support ANTXR1 as the primary cellular receptor for SVV.

We sought to determine if ANTXR1 KO cells had lost the ability to bind SVV. We also assessed the potential role of TEX2, another candidate from the HAP1 screen, in binding SVV. We incubated WT, ANTXR1 KO, and TEX2 KO H446 cells with WT SVV labeled with the fluorophore Cy5 (SVV-Cy5) and analyzed the cells by flow cytometry using the non-permissive SCLC cell line, DMS 114, as a negative control for SVV binding (Figure 22B). WT H446 incubated with SVV-Cy5 showed a high level of fluorescence (mean fluorescence (MF)=2,373) compared to DMS114 cells incubated with SVV-Cy5 (MF=425). TEX2 KO H446 cells showed a similar fluorescence profile to WT H446 cells (MF=2,233), indicating there was no loss of SVV binding ability corresponding to loss of TEX2 protein expression. In contrast, ANTXR1 KO H446 cells showed a markedly diminished fluorescence profile similar to the negative control line, DMS114 (MF=358). Loss of SVV binding was only observed in ANTXR1 KO cells, indicating that not only does ANTXR1 bind directly to
SVV based on co-IP data, but that it is the major binding determinant for the virus in intact cells.

We analyzed the complex of SVV bound to the ANTXR1-Fc chimeric protein by cryo-electron microscopy. Picornaviruses have an icosahedral capsid formed by 60 copies of a protomer consisting of three major capsid proteins VP1, VP2 and VP3 and a fourth much smaller protein, VP4 positioned in the interior of the capsid. Copies of VP1 are assembled around the five-fold axis, while VP2 and VP3 alternate around the three-fold axis (233). Our reconstruction matches the existing atomic model of the virus (101) when filtered to 18 Å resolution (Figure 22C). Additionally, the map displays the receptor subdomains distributed radially around the five-fold axis, in crown-like geometry similar to other picornaviruses, such as poliovirus (234), rhinovirus (211) or coxackieviruses(235). The map revealed the receptor binding quasi-perpendicular to the capsid close to the center of the protomer, making contact with all three major capsid proteins.
Figure 22. **ANTXR1 is the major binding determinant and cellular receptor for SVV** A. SVV-GFP was pre-incubated with the ANTXR1-Fc chimera, ANTXR2-Fc chimera, or IgG-Fc isotype control prior to a 8 h incubation with WT H446 cells. Cell nuclei were stained with a NucBlue LiveReady probe. Scale bar represents 100 µm. B. ANTXR1 KO (blue) and TEX2 KO (red) cells were incubated with SVV-Cy5 and analyzed by flow cytometry. WT H446 (black) and DMS114 (grey) cells were used as positive and negative controls for SVV binding, respectively. C. Cryo-EM density map of SVV capsid (blue) bound to ANTXR1-Fc chimera (green).
Previous studies of the ANTXR1 protein found the protein can undergo multiple post-translational modifications, including glycosylation on three possible asparagine (Asn) residues on the extracellular domain. Additionally these studies determined the interaction between ANTXR1 and the first confirmed ligand of ANTXR1, the protective antigen (PA) protein of the anthrax toxin, was dependent on the presence of the glycans on the protein. Therefore, we sought to determine whether the interaction the glycosylation of ANTXR1 was essential for the interaction between SVV and ANTXR1. To test this hypothesis, we performed a co-IP with SVV and the ANTXR1-Fc chimera after incubating the ANTXR1-Fc with the deglycosylase enzyme, PNGase F, for increasing amounts of time (Figure 23A). Proteins were eluted and analyzed by parallel protein gel silver staining (top and middle panel) and immunoblot for SVV capsid proteins (bottom panel) using the IgG-Fc chimera as a negative control. The native ANTXR1-Fc protein appears as a smear of bands of differing molecular weights due to the different glycans on individual protein molecules. Upon PNGase F digestion, the ANTXR1-Fc protein showed an initial decrease in a molecular weight and a PNGase F incubation time dependent thinning of the protein smear into a more distinct singular MW band, indicative of the loss of glycans due to PNGase F cleavage. Additionally, we observed a PNGase F incubation time dependent decrease in SVV capsid proteins on both the silver stained protein gel and immunoblot. Therefore, binding of SVV to the ANTXR1-Fc is diminished upon loss of ANTXR1 glycosylation.

Furthermore, we examined whether PNGase F digested ANTXR1-Fc proteins lost the ability to block SVV-GFP infection in H446 cells. Similar to the experiment in Figure 22 above, we incubated SVV-GFP with undigested ANTXR1-Fc, PNGase F digested ANTXR1-Fc protein, or PNGase F alone, prior to an overnight incubation with WT H446 cells and then analyzed
cells by fluorescence microscopy (Figure 23B). Cells incubated with SVV-GFP alone or with PNGase F only showed a high level of infected or GFP+ cells, confirming PNGase F does not act on SVV directly. In contrast to undigested ANTXR1-Fc that completely attenuated the SVV-GFP infection, PNGase F digested ANTXR1-Fc partially lost the ability to block SVV-GFP infection and showed a significant increase in SVV-GFP infected cells. These results confirm that similar to the interaction between ANTXR1 and the PA protein of anthrax toxin, the interaction between SVV and ANTXR1 is dependent on the presence of ANTXR1 glycosylation.
Figure 23. ANTXR1 glycosylation is important for SVV binding to ANTXR1. A. SVV was co-immunoprecipitated with an ANTXR1-Fc chimera digested with PNGase F for increasing amounts of time. Bound proteins were eluted and analyzed by silver staining (top and middle panels). Samples were analyzed by Western blot using an anti-SVV antibody in parallel (bottom panel). Undigested ANTXR1-Fc and IgG-Fc chimera proteins were silver stained as positive and negative controls, respectively. Input SVV was immunoblotted as a positive control. B. SVV-GFP was pre-incubated with undigested or PNGase F digested ANTXR1-Fc chimera, or PNGase F alone prior to a 8 h incubation with WT H446 cells. Cell nuclei were stained with a NucBlue LiveReady probe. Scale bar represents 100 µm.
Discussion

ANTXR1 functions as one of the two receptors for the *Bacillus anthracis* toxin (182). The toxin is made up of three proteins: the protective antigen (PA), lethal factor (LF) and edema factor (EF) (182). PA binds ANTXR1 and ANTXR2 to induce rapid oligomerization of the receptor-ligand complex and subsequent internalization of the tripartite complex. Previous studies have shown that PA triggers clathrin-mediated endocytosis via ubiquitination of ANTXR1, allowing for cellular entry of all proteins of the anthrax toxin (236-238). Similarly, receptor-mediated endocytosis is the most common method of viral entry into cells. Viruses, including influenza virus A and hepatitis C virus, have been previously shown to use clathrin-dependent endocytosis for cellular entry (222, 239, 240). It is currently unknown if a similar endocytotic pathway is operant in the case of SVV; however, one intriguing possibility is that the same pore-forming capacity of ANTXR1 could play a role in SVV RNA genome release.

SVV is unique among known viruses in using ANTXR1, or any related protein, as a primary receptor. In contrast to a number of previously identified receptors of other picornaviruses, ANTXR1 is not a member of the immunoglobulin superfamily (IgSF) of receptors (209). Although ANTXR1 shares common features to the IgSF receptors in being a single-pass transmembrane glycoprotein, it may be unique in its role as a receptor to both a mammalian virus and a bacterial toxin.

The ANTXR1 protein has been previously shown to be capable of undergoing multiple post-translational modifications, including glycosylation, phosphorylation, and ubiquitination (238, 241, 242). Correct protein folding and binding to PA are both dependent on the
extracellular glycosylation of ANTXR1 (241). Glycosylation, palmitoylation, and ubiquitination can all regulate cellular localization and trafficking of ANTXR1 to and from the plasma membrane (238, 241). Our current results with PNGase F confirm that the glycosylation of ANTXR1 is important for the interaction with SVV. Additionally, any of the other possible post-translational modifications may regulate the ability of ANTXR1 to interact with not only SVV, but also with other downstream proteins important in the SVV life cycle.

Interestingly, \textit{ANTXR1} was initially discovered as a tumor endothelial marker and is also known as \textit{TEM8} (243). We show that ANTXR1 is frequently expressed on the surface of tumor cells compared to normal cells. Efforts to develop a therapeutic antibody targeting ANTXR1 expressed in tumor endothelium have been hampered by cross-reactivity of the antibody to ANTXR2. The exquisite selectivity of SVV for ANTXR1 and the medium resolution cryo-EM structure described here may inform future therapeutic development in the antibody space for both anti-angiogenic purposes and a potentially novel target for select neuroendocrine cancers.

Our gene expression data suggests that beyond expression of \textit{ANTXR1}, cancer cell defects in the innate immune response are important determinants of successful SVV replication. This feature is not unique to SVV and may be a shared requirement for success of many oncolytic viruses (39). Defects in innate immune response pathways are common in cancers; SCLC in particular has been shown to frequently lack key components of MHC class I antigen presentation as well as decreased expression of immune stimulatory cytokines. Both result in decreased tumor cell recognition and removal by the immune system (244-246). These features, together with ANTXR1 expression, may define a category of cancers particularly
amenable to treatment with SVV. An improved understanding of how cellular innate immune response pathways dictate permissivity could identify synergistic combination strategies with therapeutic agents targeting these pathways in cancer cells.

The identification of the receptor for a novel virus is an important first step in understanding its biology and tropism. For SVV, this discovery is additionally germane, as *ANTXR1* expression may serve as a predictive biomarker for future clinical development, facilitating the identification of patients most likely to benefit from SVV virotherapy. Predictive biomarkers to guide future clinical trials is a particularly acute need for patients with SCLC. This is a highly aggressive and nearly universally lethal cancer, for which few tractable therapeutic targets have been identified. Many large scale clinical trials, conducted without biomarker selection, have failed to advance the standard of care for this disease (5, 247). Carefully defined markers that would focus novel therapeutic studies on the responsive subset of patients could change this field. The identification of *ANTXR1*, and possibly suppressed innate immunity, as selection criteria will help define the structure of our subsequent clinical trials.
Conclusions

SVV-001 is a novel and promising therapeutic agent for particularly lethal high-grade cancer subtypes. However, specific limitations have hindered the clinical progression of SVV-001. The studies presented here were performed in the hopes of identifying methods to overcome these limitations. The first chapter described our efforts to identify an optimized peptide for the SVV-001 encoded 3C protease for the eventual design of a specific viral protease activated peptide-prodrug. The peptide-prodrug was hypothesized as a combination therapy with SVV-001 virotherapy to increase anti-tumor efficacy prior to the production of neutralizing antibodies, which currently limit the duration of SVV-001 virotherapy. An additional and more important limitation to the clinical progression of SVV-001 was the unknown identity of the receptor for SVV-001. The second and third chapters presented here establish our work to discover and subsequently characterize the identity of the receptor for SVV-001. The novel identification of ANTXR1 as the receptor for SVV-001 is paramount in further understanding the biology and tropism of SVV-001. Furthermore, it is possible that ANTXR1 expression along with IFN response pathway expression may be used as a biomarker for SVV-001 permissivity in the clinic to ultimately identify patients that would benefit from SVV-001 virotherapy. Studies to further characterize the interaction of ANTXR1 and SVV-001 as well as ANTXR1 as a biomarker for SVV-001 permissivity in the clinic are currently ongoing.
References

17. A. Augustyn et al., ASCL1 is a lineage oncogene providing therapeutic targets for high-grade neuroendocrine lung cancers. Proc Natl Acad Sci U S A 111, 14788 (Oct 14, 2014).
46. A. Schietinger, M. Philip, R. B. Liu, K. Schreiber, H. Schreiber, Bystander killing of cancer requires the cooperation of CD4(+) and CD8(+) T cells during the effector phase. The Journal of experimental medicine 207, 2469 (Oct 25, 2010).
47. H. L. Kaufman et al., Local and distant immunity induced by intralesional vaccination with an oncolytic herpes virus encoding GM-CSF in patients with stage IIIc and IV melanoma. Annals of surgical oncology 17, 718 (Mar, 2010).
52. A. D. Buchheit et al., An oncolytic measles virus engineered to enter cells through the CD20 antigen. Mol Ther 7, 62 (Jan, 2003).


143. P. Horvath, R. Barrangou, CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167 (Jan 8, 2010).
167. I. J. Good, G. H. Toulmin, The number of new species, and the increase in population coverage, when a sample is increased *Biometrika* **43**, 45 (1956).


Curriculum Vitae
The Johns Hopkins University School of Medicine

________Linde A Miles___________  ________06/01/16_______
Name         Date of this version

Educational History:
Ph.D. expected 2016 Pharmacology and Molecular Sciences
Mentor: Charles M. Rudin M.D. Ph.D.
*Currently completing JHU Ph.D. program/thesis research
at Memorial Sloan Kettering Cancer Center

B.S. 2009 Biochemistry and Molecular Biology
Pennsylvania State University

Other professional experience:
Research rotation 2011 Duojia Pan, Ph.D., Johns Hopkins SOM
Research rotation 2010 William B. Isaacs, Ph.D., Johns Hopkins SOM
Research rotation 2010 Charles M. Rudin, MD PhD Johns Hopkins SOM

Undergraduate Research
2007 – 2010 Biochemistry and Molecular Biology/Chemistry
Pennsylvania State University
Lab of J. Martin Bollinger Ph.D.
and Carsten Krebs, Ph.D.

Scholar Awards and External funding:
2016 AACR-Women in Cancer Research Scholar Award
Memorial Sloan Kettering Cancer Center

2015-2016 Lung Cancer Research Foundation
Memorial Sloan Kettering Cancer Center

2014-2015 Lung Cancer Research Foundation
Memorial Sloan Kettering Cancer Center

Awards and honors:
Scheinberg Travel Award 2014 Memorial Sloan Kettering Cancer Center
Graduate Student Poster Award 2012 Johns Hopkins SOM; Salk Institute
Cancer Biology Training
Consortium (CABTRAC)

Oral presentations (presenting author underlined):

**Miles, L.A.**; Burga, L.; Bostina, M.; Poirier, J.T.; Rudin, C.M., Identification of the anthrax toxin receptor (ANTXR1) as the high affinity cellular receptor for Seneca Valley Virus (SVV), American Association for Cancer Research (AACR) Annual Meeting, New Orleans, LA, April 16-20, 2016


Posters (presenting author underlined):


International Meeting of Replicating Oncolytic Virus Therapeutics, June 15-18, 2013, Quebec, Canada


Publications:


between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. PNAS. DOI: 10.1073/pnas.0909649106