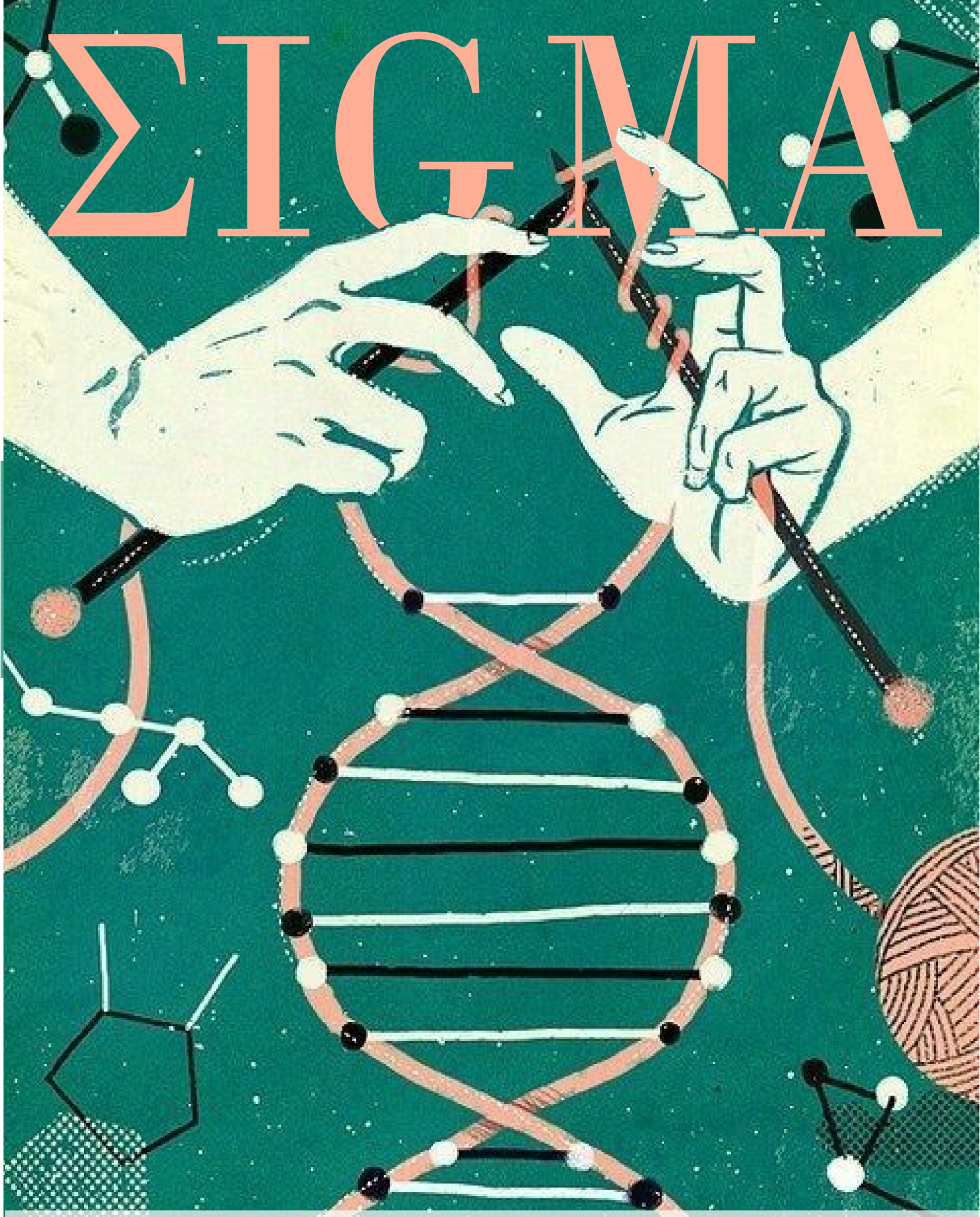
ΣIGMA



VOLUME 8 - Research Club

Welcome to the Research Club!

The Research Club is an organization that is geared towards facilitating STEM-related student research. Throughout the school year, the Research Club hosts presentations from renowned scientists, conducts biological lab experiments, prepares students for various research competitions, and holds workshops on applications to summer research programs. We are the main club at Stuyvesant that is dedicated to fostering students' research interests and preparing them for greater opportunities. Each semester, we publish SIGMA, so that members of the Stuyvesant Research Club community have access to scientific material written by their peers. We also encourage our members to annotate each paper. This serves to help the scientists of tomorrow build up their writing and communication skills that are necessary for research.

This edition features many of this year's graduating seniors' research reports, which were submitted to the Regeneron Science Talent Search Competition and various other prestigious competitions. SIGMA has expanded to include scientific research from other high schools across the city. Our third paper in this year's edition is from Park East High School. This edition also showcases a creative works project about the COVID-19 pandemic.

The club hopes to motivate its members to actively engage in scientific discussions and gain valuable experience by receiving feedback to improve their scientific writing. Throughout this spring semester, we have seen tremendous growth in all of our contributors.

Thank you for reading!

Faculty Advisors

Jason Econome

Scott Thomas

Student Leadership Board

President

Justin Lam

Vice-President

Ethan Samuel Lin

Editor-In-Chief

Neil Sarkar

Editors

Kim Firoozan, Layouts

Stephen Lee, Annotations

Leaders-In-Training

Nina Shin, Vice-President

Justin Leung, Secretary

Edward Oo, Research & Events

Kai Wei Zhang, Web Design

Arthur Liang, Finance

Editors-In-Training

Shrey Patel, Processing and Revisions

Julia Wu, Annotations

Rania Zaki, Creative Works

Contributors:

Anwesa Brahmachary

Corina Chen

Fu Chen

Judy Chen

Chris Dong

Daniel Gordon

Irene Hao

Hua Huang

Kazi Hyder

Vincent Jiang

Myat Kyaw

Isabelle Lam

Kaitlyn Lee

Amy Lin

Kevin Liu

Leonard Ma

Shivani Manimaran

Danica Moser

Christina Pan

Arian Rahman

Aryan Ruparel

Md Shad

Wasi Shahriar

Claire Shin

Liesel Wong

Grace Yang

Michelle Zhang

Thank you Principal Contreras, the Biology Department, the Chemistry Department, and the Physics Department for their support.

IN THIS ISSUE

Histology is a field of study that uses microscopes to study tissues, groups of cells with a common function

Cancer resistance is the developed resistance of cancer cells to treatment through genetic changes. In this case, the targeted treatments are resisted by the LUAD cells.

EGFR (Epidermal Growth Factor Receptor) attaches (binds) to other proteins, called ligands, outside the cell and regulates signaling pathways to control cellular proliferation.

Tumor Suppressor Genes regulate cell division. Mutations in tumor suppressor genes normally lead to cancer. The tumor suppressor genes here are RB1, which codes for retinoblastoma protein (pRB), and TP53, which codes for p53.

Metastasis is when cancerous cells spread from a primary to one or more secondary locations, resulting in increased proliferation of cancer

Down-regulation of genes reduces the quantity of the encoded protein produced, which in this case are transcription factors.

Transcription factors are proteins possessing domains that bind to the DNA of promoter or enhancer regions of specific genes.

Normally, differentiated cells are irreversibly differentiated, but some cells are plastic and can, in this scenario, differentiate into a stem cell state.

qPCR synthesizes multiple identical strands of double-stranded DNA from original DNA. In this experiment, qPCR allows for the detection of the exact amount of DNA synthesized during the PCR process by measuring the amount of fluorescence given off by SBYR Green, a DNA dye.

The Effect of RB1 Domain Deletions on NKX2-1 Expression in Lung Adenocarcinoma

Eren Ucar

Annotated by Leonard Ma, Arthur Liang, Fu Xing Chen, Hua Huang, Liesel Wong, Wasi Shahriar

Abstract

Lung cancer is one of the deadliest diseases in the world, taking more than 150,000 lives per annum in the United States. Lung cancer has three major **histologic** subtypes, with lung adenocarcinoma (LUAD) being the most common. LUAD develops in the small airway, where gases are exchanged to oxygenate blood in alveoli. Some forms of LUAD are driven by mutated growth factor receptors. “Targeted” therapies directed at these mutated receptors have been developed and are used clinically. Targeted therapies for LUADs are initially effective, but most patients acquire **cancer resistance** to them. In mutant **EGFR**-driven LUADs, one form of acquired resistance to these targeted therapies is associated with a change in cancer histology. Specifically, in EGFR-driven LUAD models, a loss of **tumor suppressor genes** *RB1* and *TP53* leads to de-differentiation of these cells, allowing them to return to a stem cell state, creating potential for **metastasis** to occur. Loss of RB1 leads to **down-regulation** of the *NKX2-1* and *GRHL2* genes which encode crucial **transcription factors** that regulate cell identity – the cells of the alveoli where LUAD originates. With a decrease in the levels of these transcription factors, these lung cancers show what is called “**lineage plasticity**”. However, histologic transformation is not well understood and no therapies exist to suppress this mode of acquired resistance. My research was focused on the mechanism by which the tumor suppressor RB1 regulates the expression of *NKX2-1*. Specifically, I investigated which domains of the RB1 protein are required to control the expression of *NKX2-1* in these EGFR-driven LUADs. To measure these effects, I generated versions of RB1 in which domains A, B, or C are deleted. To test the effect of these domain deletions on *NKX2-1*, I used **quantitative polymerase chain reaction** (qPCR) to measure differences in levels of *NKX2-1*. My experiments demonstrated that domains A and C were dispensable for *NKX2-1* expression, whereas domain B was required.

Introduction

Much is known about risk factors for the development of lung cancer. Most notably, smoking greatly increases the odds of being diagnosed with lung cancer (1).

Over the last 50 years, much insight has been gathered on the genetic and molecular mechanisms that underlie the growth of cancer. Less may be understood at the level of complex systems, such as the lung. In many instances, a cancer cell must inactivate normal tumor suppressor pathways, with the most common being mutation of the tumor suppressor gene TP53 (2). In studies of histologic transformation in the lung, tumor suppressors TP53 and RB1 are both inactivated. My project is focused on what effects follow deletion of RB1. I focused my attention on what aspects of RB1 have an effect on the expression of transcription factors that promote the growth of lung adenocarcinomas (LUADs).

One of the most common mutations observed in LUADs is in the epidermal growth factor receptor (EGFR) pathway. When EGFR is mutated to a version that stays constitutively active, it can lead to excessive growth, which is the hallmark of cancer. RB1 is a gene that is commonly mutated in many forms of cancer. Although first discovered and named after a form of pediatric eye-cancer, it is frequently mutated and/or inactivated in other cancers, including nearly 100% of **small cell lung cancers** (SCLC) and 10% of LUAD (3).

Small cell lung cancers are a fast-growing type of lung cancer that normally occurs in smokers. It usually starts in the bronchi and creates large tumors that aggressively spread throughout the body.

However, something strange happens to LUADs when they are exposed to targeted therapies when the RB1 gene is inactivated. They seem to have an ability to "change-face," or escape treatment by losing/restricting certain traits that allow them to turn into different types of cells. Specifically, in the case of these EGFR-driven LUADs, loss of RB1 plays a crucial role in this transformation. Ongoing research in the Varmus Laboratory has demonstrated that in models of LUAD in which RB1 is genetically "knocked out", there was an associated loss of **epithelial identity**. This may be due to down-regulation of key transcriptional factors responsible for epithelial identity, namely NKX2-1 and GRHL2. With a decrease in these regulators, LUAD cells are able to change face more readily and are no longer restricted to act as mature, differentiated epithelium.

Epithelial Identity is the identity or function of a specific membranous cellular tissue in the body

Importantly, previous data has shown that restoration of RB1, using a **transgene**, leads to a direct rescue of NKX2-1. This was surprising as RB1's role as a **transcriptional co-repressor** is well documented, but less is known about RB1 functioning as a transcriptional co-activator. My project was focused on determining what parts of RB1 are required for NKX2-1 expression in a model LUAD system. To address this, I chose to manipulate the known functional domains of RB1 to determine what parts of RB1 are required for NKX2-1 expression – domains A, B, and C - have been previously identified and functionally annotated.

Transgenes are genes that are artificially introduced into the genome of another organism.

Transcriptional co-activors bind to transcription factors to increase the rate of transcription.

Complementary DNA (cDNA) is DNA synthesized from a single-stranded RNA template. In this study, RB1 could be expressed as cDNA without any introns (non-coding regions).

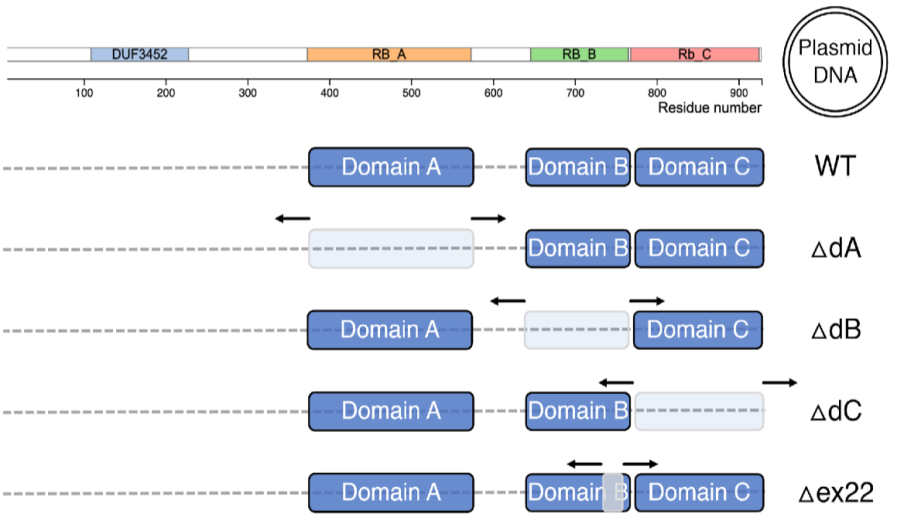
Doxycycline, a tetracycline derivative, was used to regulate transcription in the Tet-On inducible expression system. The administration of doxycycline in this inducible system allowed for the activation of the transgene expression.

To accomplish this, I first generated a synthetic version of RB1 by assembling DNA fragments into a plasmid, where the expression of the RB1 gene as an intron-free **complementary DNA (cDNA)** was inducible. The full-length version of RB1 served as a positive control, as previously demonstrated. (Supporting Information). I then split up RB1 into known annotated domains using outward facing PCR, or inverse PCR. In addition to including controls that were not treated with **doxycycline**, and thus did not express the given transgene construct, I also generated a known variant of RB1(Δ ex22) that is unstable as a mature protein, serving as a negative control for re-expression. Using qPCR, I was able to identify which specific domains of RB1 (A. B or C) were required for re-expression of NKX2-1.

Experimental Methods

Creating Domain Deletions:

1.I received a cDNA template for WT RB1 in a doxycycline-inducible, lentiviral plasmid from my mentor, Dr. Eric Gardner. I used publicly-available Uniprot™ and PhosphoSite® software to annotate the functional domains of human RB1 and designed primers around these domains, flanking the regions I wanted to remove. Through inverse PCR, I created RB1 plasmids with domain deletions termed ΔdA, Δd B, Δd C, and Δex22. A graphical representation of these constructs is provided below.



2. Following PCR, I treated the reaction with a cocktail of enzymes termed “KLD” (NEB). This is a commercial reagent that combines a kinase (K), ligase (L) and the restriction enzyme DpnI (D), to phosphorylate and ligate a linear, double stranded DNA product produced by PCR. The **methylation** specific , whereas my PCR product isDpnI enzyme helps to remove my plasmid template DNA, as plasmid DNA is methylated in the bacteria I used for sub-cloning not. I used recombination deficient, chemically competent bacteria to perform a transformation reaction (NEB Stable; NEB) and culture out my colonies.

Methylation is the addition of methyl groups to the DNA molecule to repress gene transcription.

Following overnight growth on selection plates (LB agar + carbenicillin; in house), I picked multiple colonies from each plate to culture in the presence of selection antibiotic and then isolated DNA from these cultures using a technique referred to as “mini-prep” (QIAprep Spin Miniprep kit; Qiagen). From the isolated DNA, I verified whether the sequences were correct using **Sanger sequencing** (Eton Biosciences). Additionally, I used multiple primers that covered the relevant domain regions of RB1 to confirm that the mutagenesis PCR had worked and did not introduce any unwanted DNA.

In **sanger sequencing**, the isolated DNA is replicated many times and incorporates dideoxynucleotides, which are fluorescent nucleotides that inhibit DNA polymerase to terminate the chain. This results in the creation of DNA fragments of different lengths that are run through a gel electrophoresis. The fluorescent dideoxynucleotide that ends each fragment is recorded as the fragments, from shortest to longest, cross the end of the gel, which allows for the sequencing of the DNA one nucleotide at a time.

Creating Lentivirus:

1 used an immortalized human embryonic kidney cell (HEK-293FT®) to generate functional **lentivirus** from the plasmids I cloned. To accomplish this, I transfected HEK-293FTs with our plasmid DNAs as well as with two other plasmids to help pseudotype and package the lentivirus (MD2.G and pSPAX2). Several days following transfection of these cells, the cells will release lentivirus into the culture media (DMEM + 10% fetal bovine serum; Gibco) that I collected and filtered through a 0.45um filter to remove cellular debris. After plating my cells of interest, I then infected RB1-KO cells with the various lentiviruses. One day following infection, the media on these cells was changed (RPMI + 10% fetal bovine serum; Gibco) to include a selection antibiotic (puromycin) for my lentivirus. Three days following selection, my negative control cells that were not infected with any lentivirus were killed by puromycin, whereas the cell lines that received the various domain deletion constructs were alive and could now be used for experiments.

Lentivirus is a genus of retroviruses with the capability to cause deadly diseases involving long incubation periods (i.e. the human immunodeficiency virus).

Cell Culture and Collection:

I removed my cell lines from their parental plates and seeded new 100mm plates with a sparse number of cells (~10,000/plate), the cells were counted using an automated cell counter and trypan blue as a viability indicator (Cellometer Auto T4; Nexcelom). One day after plating cells, I changed the media to include doxycycline to permit inducible expression of my transgenes, RB1 WT or the RB1 domain mutants. I changed the media every other day to +/- include doxycycline and then collected cells at 7 days following turning on the transgene. Plates of cells were washed once with cold phosphate buffered saline (PBS) + 1mM EDTA and scraped off the plates using cell scraping spatulas (Corning).

qPCR and Data Collection:

I extracted total RNA from my cell pellets using a commercial kit (RNeasy Mini Kit; Qiagen) and determined the concentration of RNA as before, using a NanoDrop device. I then used 1ug of total RNA and converted this to cDNA using a High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). Following conversion of RNA to cDNA, I then performed SYBR green-based quantitative PCR (qPCR) for my transcription factor of interest NKX2-1,

as well as an internal control used for normalization of data, the housekeeping gene beta actin (ACTB). Samples for qPCR were prepared in a 384 well optical plate and at least 5 replicates per primer pair per sample were used to capture the variability of the assay and my pipetting. Data were first normalized against ACTB values and then compared using the delta delta Ct method (4). Data were then plotted in statistical tests were performed between groups using a student's t-test (GraphPad Prism Software).

CRISPR-Cas9 is a gene-editing technology used to edit parts of the genome by removing, adding or altering sections of the DNA sequence. It uses the enzyme, Cas9 and guide RNA (gRNA) accomplish this.

Gene knockouts disable or delete certain genes, allowing us to study the gene functions by turning them on and off.

Decreased expression of NKX2-1 and GRHL2 will lead to decreased function of the epithelial cells which causes increased permeability of the alveolar membrane.

Cycle threshold(Ct) is the amount of cycles until the fluorescence generated reaches an intensity that cannot be detected anymore.

These Ct values give a relative RNA level for specific genes for sample tested. A sample with a shorter Ct time indicates that is was quicker for the machine to detect the target gene, which indicates that there is a high quantity of it.

Protein domains are independent parts of the protein chain that are responsible for a particular function in the overall role of the protein. **Exons** are segments of DNA or RNA that contain information for coding peptide sequences.

Results

The LUAD cell line NCI-H1975 used in these studies contains an EGFR driver mutation (L858R/T790M) in addition to an inactivating mutation in TP53. I used **CRISPR/Cas9 t**o delete the RB1 gene and performed RNA-sequencing across multiple **RB1 knockout**. The RNA-sequencing data suggested that in all clones where RB1 was deleted, expression of NKX2-1 and GRHL2, two genes involved in alveolar **epithelial identity was decreased**. These data suggest that RB1 may regulate the expression of the lineage transcription factor NKX2-1, a finding not previously reported.

To determine how RB1 may regulate the expression of NKX2-1, I developed versions of RB1 that lack certain functional domains. For example, domains A and C contain cyclin dependent kinase (CDK) phosphorylation sites that regulate RB1's interaction with E2F transcription factors, as well as localization and folding stability; domain B is responsible for binding of RB1 to DNA in chromatin. Using a qPCR designed to test expression the target transcription factor NKX2-1, it is possible to compare PCR **cycle thresholds (Ct)** required to amplify a given target. I used this strategy to demonstrate that as compared to WT RB1, the deletion of domains A and C did not impair the ability of RB1 to restore expression of NKX2-1, whereas deletion of **domain B** or the removal of **exon 22** did not. Taken together these data identify domain B as being required for NKX2-1 expression upon RB1 restoration.

Discussion

These data help uncouple an important role for the tumor suppressor protein RB1 from its well-known role in regulating cell cycle, by identifying a role for transcriptional regulation of lineage transcription factors in the lung (6). It is clear that domain B and ex22 deletions have the largest impact on NKX2-1 expression, as they mirror the negative control results that lacks RB1 expression. However, these data must also be considered with the understanding that the deletion of large domains may affect the overall folding, stability and localization of the truncated versions. I did not determine the stability of the domain mutants, nor did I investigate their function beyond regulation of NKX2-1, limiting a more detailed understanding of these studies.

Future work can include combining **RNA-sequencing** and **chromatin immunoprecipitation** followed by sequencing (ChIP-seq) to determine where in the genome RB1 is binding, and what other transcripts are changed in response to restoring RB1. As well, through a **Western Blot method** I could measure protein synthesis levels of these NKX2-1 models.

As well, these deletions have not been tested in **murine models** to study survival rates. This may be a worthwhile experiment in the future to test impact on mortality of these domain deletions.

RNA-sequencing is a technique that may examine the quantity as well as the sequence of RNA. Using RNA-sequencing in cancer research allows detection strand-specific information, which is very important.

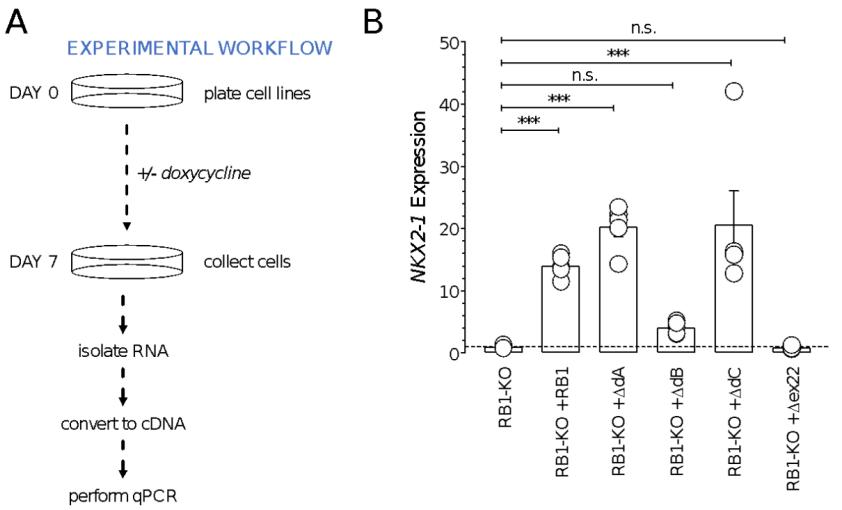
Chromatin immunoprecipitation is an experimental technique that investigates the interaction between proteins and DNA by checking DNA for binding sites of transcription factors.

Western Blot Method is an analytical technique to detect specific proteins in a sample of tissue extract.

Murine Models are experimental Murine (family of mice and rats) studies used to evaluate the ability and safety of different treatments

Supplemental Figures

Figure 1: Domain B of RB1 is required for NKX2-1 rescue



A.) Outline of experimental plan for qPCR-based determination of NKX2-1 rescue by domain deletion mutants.

B.) NKX2-1 cDNA expression normalized to the control, RB1-KO without doxycycline (far left). The stared marks indicate a significant difference in expression, while n.s indicates “not significant” by Student’s t-test, $p<0.005$ indicated by “***.” Five replicates per sample were used for this study. Error bars are shown as standard deviation from the mean.

Since the RB1's lacking domain B and ex22 have near zero NKX2-1 expression and mirror the activity of the negative control, domains B and ex22 play a role in NKX2-1 expression. The deletions of domains A and C have no effect on NKX2-1 expression.

Common Emotional Trajectories of Successful College Essays

Eric Xu

Annotated by Justin Leung, Daniel Gordon, Isabelle Lam, Liesel Wong, Michelle Zhang, Shrey Patel

Abstract

This research explored what emotional characteristics successful college essays have. By analyzing the emotional qualities of each paragraph, **emotional valence** and **arousal** were mapped out over the course of an essay. Results revealed what emotional road paths makes a college essay successful. In particular, almost all successful college essays had a Hollywood happy ending.

Emotional valence refers to the degree of the positive or negative connotation of an emotion or the intrinsic attractiveness.

Arousal refers to how calming or exciting an emotion is.

Empirical study consists of research using empirical evidence (direct or indirect observation or experience). In this case, this evidence would come from different people reading college essays.

Introduction

Text communicates informative contents and attitudinal information such as emotional states. Previous work has been done on emotional evaluations of specific genres such as movie reviews (Pang et al, 2002) and blogs (Mishne & Glance, 2006). In contrast, this paper reports an **empirical study** of emotion evaluation on college essays, a genre of text that impacts the future of millions of high school students every year.

College essays are inherently emotion-rich, because it meant to communicate a potential applicant's character and personality within a limited space. Therefore, due to intrinsic differences in each applicant's background, we expect to find a potentially diverse range of emotional expression patterns in college essays. Tracking the emotional expression along the internal structure of an essay is very revealing.

The objective of the study was to analyze how common or different emotional expressions are evolved from opening, to body, and closing among the "best" college essays. The findings of the study will help us to reveal what led those essays to be successful in terms of their emotional expression developments.

Analysis of Emotional Expression:

Work on sentiment analysis has typically focused on recognizing valence – positive or negative orientation. In addition to this measuring tool of valence, we also used arousal. In this work, we address the task of identifying expressions of emotion in text. There have been many research papers on emotion detection and machine learning.
In a work focused on learning specific emotions from text, Alm et al. (2005) has explored automatic classification of sentences in children's fairy tales according to basic emotions identified by Ekman (1992). The data used in their experiments was manually

annotated with emotion information, and is targeted for use in a text-to-speech synthesis system for expressive rendering of stories. Ultimately, rule-based emotional coding used in (Lee, et al., 2013) may be a little too robotic and not human enough. In writing, there are all sorts of sarcastic and subtle tones conveyed through text that is sometimes impossible for strict rules to measure subjectively.

Method

Research Question:

What is the overall trend of emotion expression (measured by emotional valence and arousal) in different parts (opening-body-end) of a successful college essay?

***Dataset*:**

The essays used in this study were chosen from public online websites that were considered exemplary and published within the last 5 years. All of them are within 650 words as required by Common Application. A total of 45 college essays were featured in high-profile public sources, which included New York Times (2017-2019), Harvard Crimson (2018), and Johns Hopkins Admission (2014-2019).

Essays were chosen based on the following criteria: that they were made within the past 5 years; the authors of the essays that were chosen all went to top 30 US Colleges based on the ranking from US News Report; and that essays that did not follow the **traditional Opening-Body-Closing structure** were excluded.

Emotional Type:

In this study, five emotional patterns (as shown in Table 1) were analyzed based on the changes of **emotional expression (valence or arousal)** between two consecutive parts of an essay (opening-body, body-closing). For example, if the valence of closing is greater than the valence of body, and valence of body is also greater than the valence of opening, this essay will be categorized into Valence-Upward type. If the changes are within **.25**, then it is considered the same.

Table 1: Five Emotional Pattern Types

	Opening- Body	Body - Closing
Upward	+	+
Bump	+	-
Downhill	-	-
Valley	-	+
Flatline	=	=

Dataset - 45 recent and outstanding essays were chosen for this study, all within the Common Application requirements and had the standard essay structure.

In a **standard essay structure,** openings usually consist of a chronologically-consistent anecdote, introduces the topic. The body develops the idea and the story. The closing summarizes or generalizes the topic, and reiterates the "moral" or take-away of the essay. This is important to understanding how measures of emotional valence and arousal are derived from the essays.

Emotional Expression Valence and Arousal are two popular dimensions in emotion research (Russell & Barrett, 1999; Hepach, et. al., 2011). They were coded on a Likert-scale of 1 to 9 (1=weakest, 9=strongest) in this study.

.25 is the degree of emotional change is based on the computer program used

Content Type:

Three types describing the primary essay contents were identified based on related literature (Fiske & Hammond, 2009; Staff of the Harvard Crimson (2017):

1) Achievements and Passion (AP) Many academic, athletic, and artistic achievements are mentioned. The intellectual and extracurricular achievements play a big part in this type of essays.

2) Characteristics and Identity (CI) The main topic of this type of essays is connecting with their culture, or realizing a profound truth about their character. Some essays jump and switch through different facets and aspects of themselves, or self-reflection. 3) Overcoming Obstacles and Learning From Experience (OE) Instead of expanding outwards into overarching themes, these essays focus on specific moment or a smaller set of events that helped them realize the type of person they are. They focus on a job, project that tells a lot about them.

Content Analysis

A **coding unit** is a paragraph in the essay.

First, structure and emotional expression (valence, arousal) were coded for each **coding unit** - a paragraph. Second, based on the differences of emotional expression among three parts of the essay, emotional types were calculated. Finally, each essay was assigned a content type (AP, CI, OE).

Two coders coded a sample of 10 essays based on the coding manual that was developed for this study until the inter-coder reliability reached an acceptable level (Cohen's Kappa=.74 for valence, .87 for arousal). Two coders reached a complete agreement on coding of structure and content type. After that, one coder coded the rest of the 35 essays.

Results

Table 2 showed that our sample essays were equally distributed among three different types of contents: AP, CI, OE.

Table 2: Number of College Essays by Content Types

Essay Type	Number of Essays	Percentage
AP	15	33%
CI	16	36%
OE	14	31%

Table 3 indicated that emotional valence patterns and emotional arousal patterns showed some similarities, which is the majority of our sample essays showed Upward, Valley, Bump, and Flatline. Downhill did not occur in valence type and only 2 cases in arousal type.

Table 3: Number and Percentage of College Essays by Emotional Patterns

Emotional	Number of	Valence Percent	Number of	Arousal Percent
Bump	6	13%	12	27%
Valley	16	36%	16	36%
Upward	19	42%	11	24%
Downhill	0	0%	2	4%
Flatline	4	9%	4	9%

In the following, we report our findings on whether there are differences in terms of emotional expression over development of the story (structure), and how emotional expression varies among three different types of contents: AP, CI, OE.

ANOVA of the effects of structure and content type on emotional expression is analyzed.

- Main effect of content type on Valence is significant ($F(2,42)=3.50$, $p<.05$), but on Arousal is not significant ($F(2,42)=.05$, $p>.05$). The post-hoc Tukey analysis showed that emotional valence of AP essays are significantly higher than that of OE essays ($p<.05$).
- Main effect of structure on both emotional valence and arousal is significant (valence: $F(2,84)=20.88$, $p<.01$; arousal: $F(2,84)=5.65$, $p<.01$). The post-hoc Tukey analysis showed that essay closings showed significantly higher valence and arousal than essay opening and body ($p<.01$).
- Interaction effect of content and structure is only significant on arousal ($F(2,84)=.99$, $p<.01$).

ANOVA (Analysis Of Variance) is a group of statistical models along with their associated estimation procedures, which is utilized to study the differences between group means within a sample.

Even though AP essays have more emotional valence than OE essays, after analyzing the effects of valence and arousal types and structure on emotional expression, all three types of essays exhibit much higher valence and arousal in their closing, similar to classic Hollywood movie happy ending.

AP essays have a higher emotional valence when compared to OE essays. Additionally, the essays consistently close with higher valence and arousal than they open with.

Conclusion

Ultimately, many successful college essays have on average a slightly positive valence. This only makes sense, as most students want to paint a positive image of themselves so they seem a better candidate for admission. Similarly, the arousal was also slightly higher than the middle point of 5 on a 9-**likert scale**, most likely because college essays ask students to condense a lot of subject material in a limited amount of space, so students readily use emotion to emphasize specific qualities about themselves.

Likert Scale - a psychometric (psychological measurement) scale where respondents specify their level of agreement or disagreement on a symmetric agree-disagree scale for a series of statements.

Other trends include significant correlations between the specific content of college essays and specific structural elements. This suggests these “combinations” of essays may be effective and overrepresented in the pool of successful college essays.

This study chose a college essay, a topic most relevant to the author’s current life to study the emotional expression analysis of communication. Future research may expand on emotional journeys in literature, and could apply our methodology to analyze the emotion of other genres such as mystery and romance novels. In addition, machine learning techniques can be used to automatically process essays in accordance to our coding schema.

The Relationship Between GPA and Sleep Hours on Weekdays

Fanta Doumbia

Annotated by Wasi Shahriar, Kevin Liu, Danica Moser, Michelle Zhang

Abstract

This paper discusses an experiment with students including their sleep habits and their grade point averages. If they get more sleep, will their GPA increase? If they get less sleep, will their GPA decrease? In this experiment, neither of these hypotheses were correct. This experiment was conducted in Park East High School. Forty surveys were printed then given out when students were free. After the data was collected, it was divided into two sections; a pile to keep and a pile to disregard because the answers did not make sense. The data was entered in a chart and a scatterplot so that the significance could be correlated. There proved to be no correlation between these two variables.

Forty students were surveyed on their sleep hours and GPAs to find out if there was a correlation between them. Some surveys weren't counted since participants misunderstood the survey, and using inaccurate data would skew the results of an analysis.

Introduction

This topic is important because it discusses the effect sleep has on performance in school, which could affect the future. For instance, if one sleeps late, they are more likely to fail in the future because fatigue interferes with getting work done. If one is getting enough sleep, they are more likely to be alert in class, participate and do all their work on time. One might also earn higher grades because they are able to focus, instead of being tired and off task.

In a previous study, 222 students (152 women, 70 men) were surveyed about sleep disturbance, the amount of sleep they got, and the scores they received on a **"SAW" (Sleep Disturbance Ascribed to Worry Scale) test**. It turned out that there was not a significant relationship between sleep disturbance ascribed to worry and sleep length and scores (Kelly, 2002). In another previous study, researchers had picked two countries to study and selected participants using **stratified cluster sampling**. They studied the students' bedtime and wake up time, including weekends. The results revealed that there was no significant relationship between sleep and conduct problems (Lin & Yi, 2014).

The **Sleep Disturbance Ascribed to Worry Scale test** is a series of questions that relates sleep to worry, such as "How often are you unable to stop worrying at bedtime". In this study, the test was used to show if there's a relationship between the sleep length and the score they got on the test.

Stratified cluster sampling is a method of collecting sampling participants by splitting the total population into different groups and then divided up into clusters. Then a random number of clusters will be chosen as participants. This method ensures that the resulting samples represented the population of students in the two counties where the study was conducted.

In another study by Wiggins and Freeman (2014), high school students in grades 9 through 12 were surveyed about school and work experiences based on gender, ethnicity, and type of work. Work hours turned out to have a negative effect on students' self-reported grades. There was no significant relationship between the number of work hours and grades. In another study by Singh (2007), researchers studied different factors that increase or decrease students' GPA. Employment was one factor that was

considered. Students who had jobs sometimes chose easier classes so that they would not have to worry about their grades. Some people prioritized their jobs over their school work. It turned out that there was a positive relationship between being employed and having a high GPA because students who had jobs took easier classes, which resulted in higher grades. The present study will investigate a different population from the studies in the previously published articles, using data mainly from two specific grades. The research question was: Is there a relationship between number of hours of sleep and GPA. The hypothesis was that there would be a positive correlation: as the number of hours of sleep increases, the GPA increases. This could be because when one gets more sleep, they are well rested and able to stay awake in class. When one is well rested, they are able to participate more (*Heart Disease Weekly*, 2005).

Results

The hypothesis was that as hours of sleep on weekdays increase, there will be a decrease in GPA. In this investigation, the significance test used was a correlation. It was established that there was no correlation between the two variables **r(30) = -.085, *p* > .05.** The scatterplot can be found in Appendix B.

The **r-value** is the correlation coefficient. It measures the strength and direction of a linear relationship between two variables.

The **p-value** helps you determine the significance of your results.

Discussion

Results in the study of hours of sleep and GPA showed no correlation between the two variables. This means that the hypothesis was not supported, because the hypothesis stated that if the number of hours of sleep increases, then the GPA increases.

A possible theory is that some students stayed up late to complete their work, getting less sleep and higher grades than the ones who got more sleep. Another possible factor, such as work, affected the students' performance in school. The students that had jobs sometimes prioritized their work over schoolwork because they were getting paid (Wiggins & Freeman, 2014). Additional possible factors were the students' health and productivity.

The results of this study show that there was no apparent correlation between the amount of sleep and GPA. This may be due to other factors such as work, productivity, and health.

One **problem** faced during the present study was that there were misinterpretations of the survey question, "How many hours do you sleep per week-day, including naps?" Participants thought that they were asked how many hours of sleep they receive in a week, when really the question asked was how many hours of sleep they get in a day. There was also possible bias due to the fact that specific grade levels were surveyed instead of all grades. Most of the subjects were seniors and freshmen. In the future, researchers could improve the present study by being more specific to prevent absurd answers that would lead to the omission of certain surveys. Further, all grades instead of just two could be included and phrase the question to "How many hours of sleep do you receive in a day, including naps?" so that participants are able to give more reasonable responses.

Problems such as misunderstood questions and different grades could have altered the data, leading to the conclusion that there is no correlation.

The questions on the survey were both nominal and quantitative. In total, there were five questions: two primary questions and three distractor questions. Questions 1 and 2 were used to evaluate the research question, while questions 3, 4, and 5 were distractor questions.

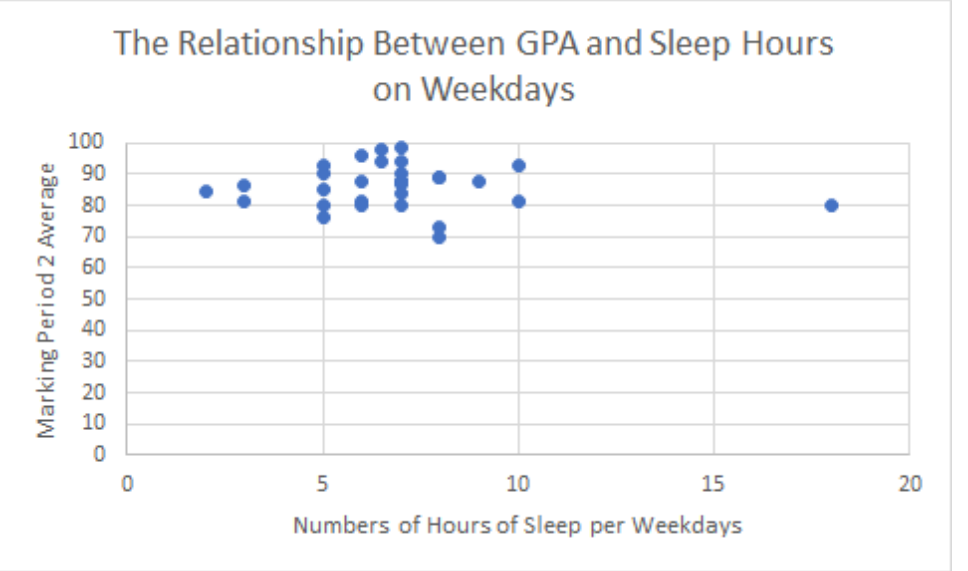
There were 40 people surveyed. Eight surveys were omitted. One was omitted because the question was misunderstood by the participant, so they provided an unreasonable answer. Another was omitted because 0 was recorded for the number hours of sleep, and the other six were omitted because they listed over 24 hours of sleep in a day. As a result, 32 participants were actually included in the sample. It's important to report why surveys were omitted and identify factors that could skew the results of the study otherwise.

The scatterplot shows that it's possible for people getting any range for sleep to have around the same GPA. This data doesn't suggest that more hours of sleep would result in a higher GPA, or that less hours of sleep would necessarily result in a lower GPA which disproves the hypotheses. This result would bring us to our discussion on what the experiment failed to do and what other factors would affect GPA besides sleep.

Appendix A: Survey

1. How many hours do you sleep per week-day, including naps?
2. What was your MP2 report card average?
3. What is your favorite ice cream flavor?
4. What is your favorite sport?
5. What do you do for fun?

Appendix B: Scatterplot



DNA Barcode Library as a tool to increase the screening throughput of human induced pluripotent stem cells and their derivatives

Tyseen Murad

Annotated by Amy Lin, Fu Xing Chen, Hua Huang, Kevin Liu, Danica Moser, Arian Rahman, Liesel Wong

Abstract

Chemical screening approaches are constantly being developed and are now used to answer key questions in developmental biology and disease treatment. However, the screening platforms currently available still produce rather low **throughputs**. This study focuses on developing a **lentiviral DNA barcode library** as a potential tool to increase the throughput of chemical screenings based on **human induced pluripotent stem cells (hiPSCs)** and their cell derivatives. DNA barcoding enables scientists to track individual cells and quantitatively assess their clonal contributions over time9,10. With this study we seek to create a platform based on genetic barcoding applicable to stem cells. To do so, we generated a DNA barcode library of 100 different lentiviral barcoded transfer **plasmids** and we **transduced** different hiPSC lines, each with a specific barcode from this library. These cell lines will later be pooled together and treated with a stimulus from a library of chemical compounds, hormones, growth factors, or a virus. The main advantage of this strategy is the increased throughput of resulting hiPSCs-based chemical screenings, enabling for more cost-effective and less time-consuming assays, as many experiments can be performed in parallel. This approach may also enhance experimental precision and accuracy. Our results emphasize the increasing importance and contribution of genetic barcoding in providing quantitative data in demanding fields such as stem cell biology.

Introduction

Stem cell biology is a rapidly developing field that combines the efforts of cell biologists, geneticists, and clinicians to better understand the genetic programs and molecular mechanisms regulating **cell fate acquisition** and how the disruption of these mechanisms results in pathological conditions.

Within this endeavor, high-throughput chemical screenings are an emerging approach to identify the small molecules controlling stem cell differentiation and reprogramming. Moreover, hiPSCs-derived cells and organoids used to model

Throughputs is the rate of production or rate something is processed.

Lentivirus is a genus of retroviruses. One example of a lentivirus would be HIV. They are one of the most efficient methods of gene delivery because can integrate a lot of viral cDNA into the DNA of the host cell.

DNA barcode library - Using libraries of diverse sequences in lentiviral vectors, each cell in a population of a million can be labeled with a unique nucleotide sequence, aka a barcode.

Human-induced pluripotent stem cell (hiPSCs) is an immature cell that is generated from adult (mature) cell and that has regained the capacity to differentiate into any type of cell in the body.

A **plasmid** is a small, DNA molecule within a cell that can replicate independently of the chromosomes.

Transduction is the transfer of genetic material from one microorganism to another by a viral agent. The 293T cells with transfer plasmids are the viral agents used to insert foreign DNA into the hiPSC lines.

Cell fate acquisition, or cell fate determination, is how a cell will develop into a final cell type. The processes that develop the cell into its final form include differentiation, cellular movement, the cell cycle, and programmed cell death.

In vitro refers to how the hiPSCs-derived cells and organoids modeled in a test tube, culture dish, or other contained area instead of in an open living environment.

PCR is a technique used by biologists to create multiple identical copies of double-stranded DNA (dsDNA) from one original dsDNA strand. In this study, there are 100 barcodes, so there will be multiple copies of these barcodes after PCR.

Vectors are DNA molecules that are used as a vehicle to artificially carry foreign genetic material into another cell.

293T Cells refer to Lentiviral Packaging and Infection in the Methods and materials section.

NGS is technology that is able to sequence more than 1 million base pairs. It can also detect abnormalities with less DNA than needed for Sanger sequencing. Some NGS types are Illumina sequencing, Roche 454 sequencing, and Proton/PGM sequencing.

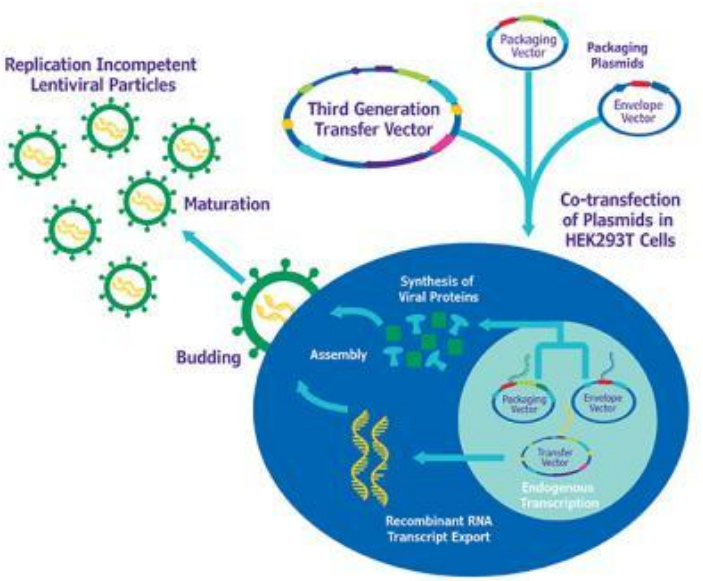
Cell-autonomous is a genetic trait in multicellular organisms in which only genotypically mutant cells exhibit the mutant phenotype. Conversely, a nonautonomous trait is one in which genotypically mutant cells cause other cells to exhibit a mutant phenotype.

Committed stem cells are in the stage in their development when the number and type of cells they can differentiate into is limited.

in vitro human diseases can be employed in high-throughput chemical screenings for drug testing. Currently, the throughput of hiPSCs-based screenings is still relatively low and can be optimized1. The purpose of this study is to develop a platform to increase the throughput of the chemical screenings using genetic barcode technology applied to stem cells.

To create the DNA barcode library, 100 barcodes that are 12 base pairs (bp) long underwent **PCR** and cloning into a lentiviral transfer **vector.** Once these lentiviral barcoded transfer plasmids were created, they were packaged into **293T cells** (Figure 2) in order to create the viruses which later were used in the transduction of each hiPSC line. These cell lines were later pooled together and treated with a stimulus. The genomic DNA was then isolated, amplified by PCR, and then sent for **next-generation sequencing (NGS)**. The main advantage of the developed platform is that it may increase the throughput of chemical screenings and enhance their precision as pooled experiments minimize batch to batch differences. Furthermore, this method provides an initial look into whether the cell mechanisms that mediate the cell response to the administered stimuli are **cell autonomous** or cell non-autonomous. The developed platform has a wide range of applications, including infectious studies, drug screenings, transplantation studies, and research focusing on cell line-specific biases in **lineage commitment**. Overall, the developed method provides a valuable strategy to increase chemical screening throughput of hiPSCs and their derivatives, furthering stem cell technology and medicine.

Figure 2: This is a breakdown of the transfection of plasmids in 293T cells, the first half of lentiviral packaging. Once the 293T cells are infected, they mass produce the virus that contains the genetic barcode which is going to be used in the transduction of hiPSC lines.



Methods and Materials:

1. Generation of Barcoded Lentiviral Plasmids

A PCR based cloning approach was used to produce the lentiviral transfer plasmid. We decided to use this cloning technique as it is very adaptable and enables the insert of nearly any DNA sequence into a backbone vector of choice. In this study, Lenti-Cas9-2A-Blast plasmid (Addgene, #73310) was used as the lentiviral backbone vector.

The amplification of the insert through PCR is important in order to produce a sufficient amount of insert for cloning. The PCR products, corresponding to the sequences of the barcode plus the **fluorescent protein**, and the lentiviral backbone vector Lenti-Cas9-2A-Blast plasmid (Addgene, #73310), were digested with the restriction enzymes. **Ligation** between the lentiviral backbone vector and the insert was performed using T4 DNA Ligase (NEB, #M0202).

Transformation was performed using **Stbl3 chemically competent cells** (ThermoFisher Scientific, #C7373-03). One Shot Stbl3 Chemically Competent E. coli are designed especially for cloning plasmids that contain direct repeats. Unlike other bacterial strains, these cells reduce the frequency of homologous recombination between long terminal repeats (LTRs) found in lentiviral plasmids and other retroviral vectors.

After transformed bacteria had time to proliferate, multiple colonies from each batch were picked and amplified in 5mL **Luria Bertani (LB) broth** supplemented with carbenicillin selection.

Fluorescent proteins are convenient markers for gene expression because it exhibits fluorescence. In this study, they were used to see if the transfection had occurred properly

Ligation refers to the joining of two molecules, in this case, the lentiviral backbone vector and the insert.

Stbl3 chemically competent cells are designed for chemical transformations of DNA containing long terminal repeats found in lentiviral vectors. One Shot Stbl3 Chemically Competent E. coli are designed especially for cloning plasmids that contain direct repeats. Unlike other bacterial strains, these cells reduce the frequency of homologous recombination between long terminal repeats (LTRs) found in lentiviral plasmids and other retroviral vectors.

Luria Bertani (LB) Broth is a nutritionally rich medium primarily used for the growth of bacteria.

2. Plasmid Quality Controls and Amplification Plasmids

After the DNA was isolated, it was quantified in order to calculate purity and concentration. Samples that showed the expected pattern of bands were sent for Sanger sequencing at Genewiz in order to confirm that the correct insert has been inserted into the new plasmid, to assess the barcode sequences obtained, and to rule out that recombination has occurred. The reactions that showed correct sequences corresponding to new barcodes from the Sanger Sequencing were further amplified for lentiviral packaging.

3. Lentiviral Packaging and Infection Amplification Plasmids

For lentiviral packaging, we performed a **calcium phosphate transfection**, for which 293T cells were used. 293T is a human cell line, derived from the human embryonic kidney 293 (HEK 293)

Calcium phosphate transfection - Calcium phosphate facilitates the binding of the DNA to the cell surface. DNA then enters the cell by endocytosis. Here it's used for lentiviral packaging in order to introduce lentivirus to the cells.

cells, that express a mutant version of the SV40 large T antigen. These cells are frequently used in biology for lentiviral packaging, which is the purpose that they serve in this study. A spin infection protocol was followed to transduce the library of hiPSCs with the barcoded lentiviruses.

Results/ Discussion

We created a lentiviral DNA barcode library and infected different hiPSC lines with it in order to be able to increase the throughput of chemical screenings. Within the limited time of the present study, we successfully created the DNA barcode library comprised of 100 lentiviral transfer plasmids which was first checked through Sanger Sequencing.We have sent the whole library of 100 different barcoded lentiviral transfer plasmids for Next Generation Sequencing (NGS) and we are currently waiting for the results, which will validate that the sequence of each plasmid is correct.

Packaging plasmids and **envelope plasmids** are plasmids that encode virion proteins. Packaging plasmid psPAX2 and envelope plasmid pMDG2G were put into 293T cells so they could monitor fluorescence expression.

We co-transfected each barcoded transfer plasmid with the **packaging plasmid psPAX2** and the **envelope plasmid pMD2G** in 293T cells for lentiviral packaging. Successful lentiviral packaging was first assessed by monitoring fluorescence expression in **transfected** 293T cells, as shown in Figure 5.

Transfection is the process of bringing nucleic acids into eukaryotic cells.

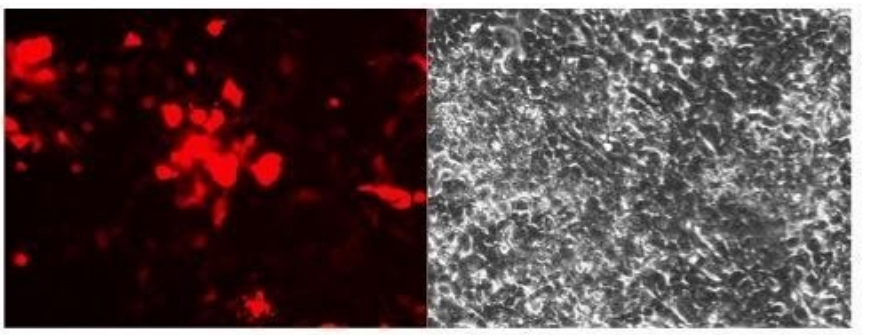


Figure 5: Representative images acquired at 20X magnification of a 293T cell culture transfected with the transfer, packaging, and envelope plasmids for lentiviral packaging, 48 hours post transfection. On the left, a representative image showing mCherry fluorescence which was monitored to assess that the transfection occurred properly. On the right is reported a bright field image of the same field of view to provide an idea of the cell density.

A **titer** is the concentration of an antibody, as determined by the finding the highest dilution at which it is still able to cause the antigen to clump (by titration).

Though we created the 100 transfer plasmids and produced the respective lentiviruses, we have not yet finished **titering** all the lentiviruses and transducing the different hiPSC lines. In the future, these lines will be pooled together and infected with different stimuli and more work will be done to validate the reliability of this platform.

Conclusions

Although the chemical screenings applied to hiPSCs and their cell derivatives are a powerful tool for controlling stem cell differentiation fates and for identifying promising drugs to treat specific disease conditions, their throughput can be increased.

The ultimate goal of our study was to be able to create a platform using barcode technology in order to increase this throughput of chemical screenings. By **increasing throughput**, we are ultimately making drug screenings in a pharmacological context, much more cost-effective as many more experiments can be performed in parallel to each other. In addition, experiments that would not be performed otherwise due to their elevated costs will be more readily available to users. Being able to perform these experiments at the same time and in the same experimental conditions will also reduce batch to **batch effects** which are often an issue.

Increasing throughput - This study was administered to try to increase the throughput of chemical screenings to make them cheaper and more efficient.

A **batch effect** is when a non-biological factor causes a change in the data produced by the experiment. This makes the data inaccurate since another variable was added to the experiment.

This platform can also find application in many different aspects of basic research in developmental biology. Overall, this platform can enable us to assess the effects of different stimuli on hiPSCs. Such stimuli can include but are not limited to viruses, chemical compounds as seen in drug applications, hormones, and growth factors.

However, much of this research is still on-going and additional work still needs to be done to validate this platform. Potentially, our findings may be able to increase throughput of chemical screenings based on hiPSCs and their cell derivatives, leading to a more efficient way of performing drug and disease experiments.

The impact of histone modification H3K9M on genome integrity within PTEN-deficient breast cancer cells

Janen Khan

Annotated by Daniel Gordon, Kevin Liu, Leonard Ma, Arian Rahman, Liesel Wong, Grace Yang

Tumor Suppressors are genes regulating cells during cell replication and division. They are also referred to as an anti-oncogene.

Histones are proteins found in chromatins, and provide structural support for chromosomes. Partly due to DNA being wrapped around histone, they take part in gene regulation. They also take part in the production of chromosomes.

Radiotherapy PTEN-deficient cells refers to radiotherapy treatment commonly used for breast cancers. It is suggested that these cells may demonstrate resistance to such radiation. The PTEN-deficient cells are cancer cells, as they will be for the rest of this paper.

Histone modification is a post-translational modification (PTM) to histones. H3K9M is a PTM that stops the widespread methylation of histones.

Methylation refers to the addition of or to substitute in methyl group(s) to a molecule or compound. It is because H3K9M can mimic the inability to methylate, that we can focus on the effects of histone modifications.

The **P13K pathway** is an intracellular signal transduction pathway that promotes metabolism, cell survival, growth, and angiogenesis (development of new blood vessels) in response to extracellular signals. This pathway can also promote the development of tumor(s).

Transcription is the first of several steps of DNA in which a particular segment of DNA is copied into RNA by the enzyme RNA polymerase.

Abstract

Cancer is quickly becoming one of the major causes of death worldwide. PTEN (phosphatase and tensin homolog) is a **tumor suppressor** that has been found to be frequently deleted in a range of cancers. The loss of PTEN primarily results in cells exhibiting cell cycle deregulation and fate reprogramming. I conducted a series of in-vitro experiments to verify another suspected relationship between PTEN and **histone** modifications in cancer cells: specifically observing the effects of histone modifications on moderating DNA damage repair in cells lacking PTEN function. My results suggest that histone modifications can increase resistance to **radiotherapy PTEN-deficient cells**. Further study of my findings can lead to groundbreaking discoveries in tumor suppressor activity and transform the way that cancer therapy is designed and delivered.

Introduction

PTEN mutations are increasingly being linked with cancer while the tumor suppressor gene's sphere of functions is still poorly understood. The most prevalent cancers today include cancers that are affiliated with high levels of PTEN mutations, such as colorectal and prostate cancer. This study expands on the effects of a common **histone modification**, specifically H3K9M (which mimics **methylation**), on DNA damage and repair systems. Due to its importance in inhibiting the **oncogenic PI3K pathway**, PTEN has been labeled as an authentic tumor suppressor (Khan, 2013). Soon after the identification of PTEN as an antagonist of PI3K signaling, multiple studies have revealed its function in regulating cell cycle progression, metabolism, and epigenetic pathways (Maehama et al., 2001). PTEN plays an essential role in activating the DNA damage checkpoint to prevent genetic instability.Moreover, PTEN plays a vital role in maintaining the condensed state of the chromatin structure with H3K9 methylation. Specifically, H3K9me3 binds heterochromatin protein 1 (HP1) to constitutive heterochromatin. HP1 is responsible for **transcriptional** repression and the actual formation and maintenance of heterochromatin. Loss of PTEN results in the disruption of the complex formed by histone H1 and heterochromatin protein HP1a (S.-Q. H

ou et al). Additionally, the inability to condense suggests active transcription at **heterochromatin** sites and can contribute to **genomic instability** through uncontrolled gene activation. Thus, in the absence of PTEN, cells cannot participate in chromatin condensation and have flaws in cell division.

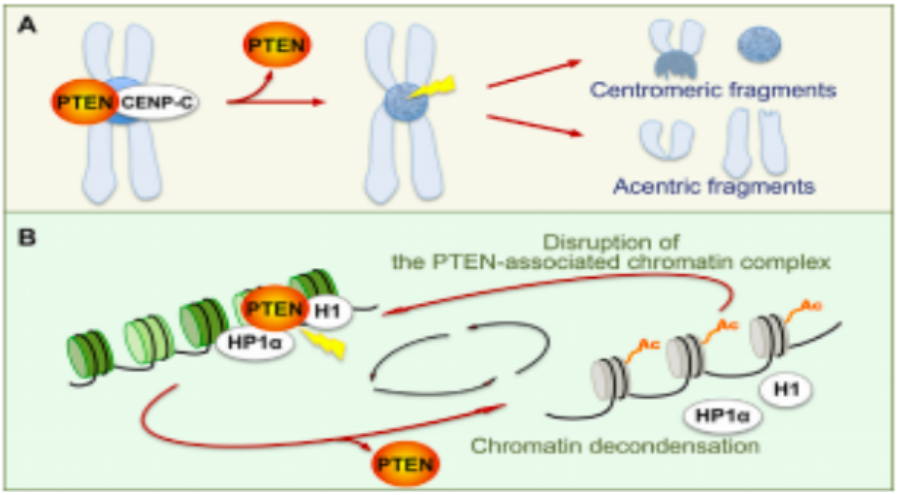


Figure 1: PTEN aids the condensation of heterochromatin by binding with H1

PTEN controls the mitotic machinery by maintaining normal architecture of the mitotic spindle and promoting high-fidelity chromosome alignment and segregation. In the absence of PTEN, uncontrolled **phosphorylation** of AKT leads to centrosome amplification and spindle shortening, spindle pole fragmentation, chromosome misalignment and non-disjunction, polyploidy, as well as mitotic catastrophe [S.-Q. Hou et al.].
H3K9M enhances the interaction of histone H3 tail with the H3K9 **methyltransferase** Clr4 and blocks the methylation of corresponding lysines on wild type histones (NCBI). By mimicking the inability to methylate, studying **H3K9M** serves as a method of isolating issues caused by the lack of PTEN and focusing on the effects of histone modifications. Specifically, by observing **double-strand breaks** and comparing the cells' ability to repair, I studied whether or not modifications were further adding instability to the genome and making it difficult to kill tumor cells.

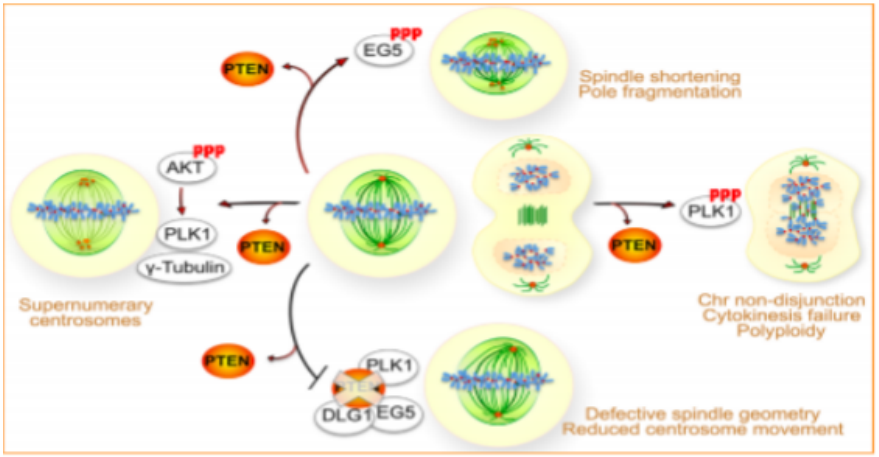


Figure 2: PTEN plays a significant role in ensuring proper cell division

Heterochromatin is tightly packed chromosome material

Genomic instability means a high frequency of mutations. These mutations can include changes in nucleic acid sequences, chromosomal rearrangements or aneuploidy (an abnormal number of chromosomes).

PTEN acts as a tumor suppressor gene through the action of its phosphatase protein product. Phosphatase is involved in the regulation of the cell cycle, preventing cells from growing and dividing too rapidly.

Phosphorylation is the attachment of a phosphoryl group. It is especially important for protein function and allows cells to accumulate sugars.

Histone methyltransferases are a group of enzymes that catalyze the transfer of methyl groups to the lysine or arsine amino acids on H3 and H4 histone proteins.

H3K9M's ability to mimic methylation makes it a key factor in studying dilemmas caused by PTEN deficiency.

A **double-strand DNA break** (DSB) occurs when both strands of the DNA duplex are severed.

Results

DNA Repair:

Figure 1:
Cells were divided into four different groups for observance:
- C6 Negative Control (dysfunctional PTEN)
- C6 Mutant H3K9M (dysfunctional PTEN and H3K9 mutation)
- WT Negative Control (functional PTEN and H3K9)
- WT Mutant H3K9M (dysfunctional H3K9)

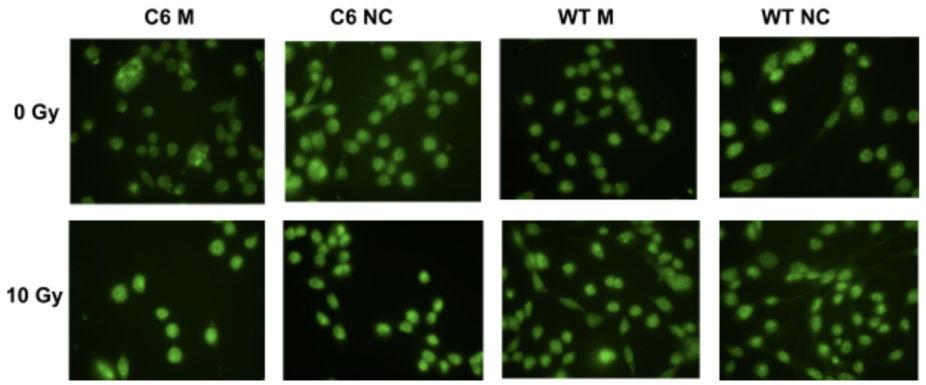


Figure 1: *Cells observed under* ***Rad-51*** *staining*

The **RAD-51** protein family which assists in repair of DNA double strand breaks; as a result, the number of RAD-51 foci corresponds to the amount of DNA repair.

A **prognosis** is a forecast of the likely course of a disease.

C6 cells were derived from a metastasizing mouse breast cancer cell line (TS/A) and were made into PTEN knockout cells.

WT Cells are an abbreviation for Wild Type. Wild Type refers to the typical form of a species as it occurs in nature.

0Gy and 10Gy radiation were used to treat the cells to measure DNA damage and repair mechanisms

Transfection is infection of a cell with free nucleic acid.

γ-H2AX is the phosphorylated form of the histone protein H2AFX; it forms when double-strand breaks appear. As a result, the number of γH2AX foci per nucleus is an accepted measure of the number of DNA double-strand breaks in single cells.

A greater number of foci indicate more repair mechanisms are present and helping the cell return to its normal state; in this case, the cells would return to their cancerous state. A lower number of foci indicate less DNA damage repair and poor **prognosis** in returning to the normal state; in this case; the cancer cells would not be able to fix themselves. As seen in Figure 3, the number of foci visible in **C6 cells** is significantly higher compared to **WT cells** in both grades of radiation. Rad-51 repair remains consistent in both **0Gy and 10Gy** in WT cells without the H3K9M mutation. The number of foci in C6 cells without the H3K9M mutation and WT cells with the H3K9M mutation are relatively similar regardless of treatment. Interestingly, Rad-51 repair seems to be most abundant in cells lacking PTEN and **transfected** with H3K9M after 10Gy radiation.

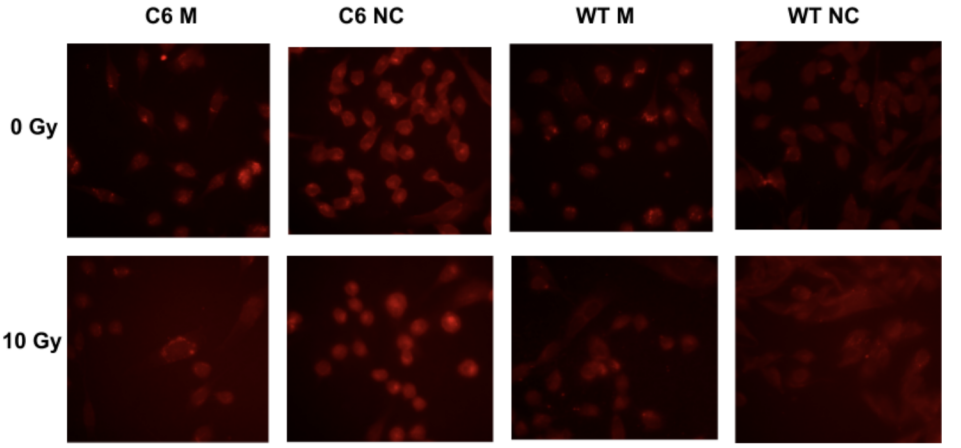


Figure 2: *Cells observed under* ***γ-H2AX***

A greater number of double strand breaks imply added genomic instability and aberrant cell cycle progression; in this case, the cancer cells would be more damaged with more double strand breaks. A lower number of double strand breaks point to greater chances of cell survival; in this case, the cancer cells would be less damaged with lower numbers of double strand breaks. In Figure 4, the number of DNA double-strand breaks is significantly higher in WT cells compared to cancer cells. As for WT cells with histone modifications and those that lack PTEN, they show relatively similar numbers of foci for double strand breakage. The C6 cells that lack PTEN and have histone mutation H3K9M have less foci stained and thus express lower numbers of double strand breaks.

In essence, modified histones in PTEN cells could cause greater resistance to radiation. This is supported by the lower number of severed DNA in C6 mutant cells.

Discussion

The immunofluorescence assay assessed mitotic errors and permitted the observation of the number of cells undergoing the cell cycle. The key findings of my study indicate that histone modifications may be causing PTEN knockout cells to exhibit greater resistance to radiation. Hypothetically, H3 should not be able to bind to the heterochromatin without the presence of methylated H3K9. Since unmethylated H3K9 disables heterochromatin folding and PTEN knockout also leads to disrupted folding, the similar number of foci for C6 mutant cells under both DNA damage and repair is valid.

The greater number of Rad-51 foci in C6 mutant cells suggested that the cancer cells have a greater ability to repair themselves and thus can proliferate even more then merely PTEN knockout cells. The lower number of double strand breaks in C6 mutant cells implies that they confer greater resistivity to radiation and thus require additional mechanisms for destruction. The added instability of histone modification H3K9M, which causes an overdrive in transcription because the heterochromatin remains unfolded, seems to be fortifying the cancer cells and promoting their survival.

A **western blot** is a laboratory method used to detect specific protein molecules from a mixture of proteins.

A future experiment would be to further study PTEN function by observing if H3 would still bind to the chromatin without PTEN present.

My next steps would involve separating the heterochromatin from my cells and doing a **western blot** to see if H3 can still bind to the chromatin without PTEN present. This would expand the understanding of PTEN function in relation to histone modifications. Since histone modification H3K9M seems to be causing genomic instability in a similar fashion to a deficiency in PTEN, the experiment proved that the presence of histone modifications provide breeding grounds for tumor cell proliferation and survival. Furthermore, the **inability of radiation** to destroy double strands and impair repair mechanisms implies the need of treatments besides radiotherapy for PTEN deficient cancer cells. Further study of my findings can lead to groundbreaking discoveries in tumor suppressor activity and transform the way that cancer therapy is designed and delivered.

This study suggests that since **radiation was unable** to destroy all cancer cells, another approach besides radiotherapy should be researched for PTEN deficient cancer cells.

As an RNA virus, **the dengue virus** has a very high mutation rate that allows it to evolve rapidly. When viruses co-infect the same cell, they are also able to swap genetic information, which is RNA in the case of the dengue virus, through a process known as recombination. These processes of mutations and recombination produce genetic variation which allows for the evolution of viruses through natural selection.

A **serotype** is a strain of a microorganism that is distinguished by the cell surface antigens that they have. They can be distinguished serologically by examining a blood sample for the presence of antibodies against the microorganism. A person's developed immunity to a single species of a pathogen is serotype-specific as antibodies against the serotype's distinct antigens are produced.

Hamming distance is a metric for comparing two data strings and is the number of amino acids that are different in a specific sequence compared to the parent sequence. It is the number of substitutions needed to change one protein string into the other. It essentially acts as the fitness attribute, which is used to summarize how close the sequence is to the parent sequence.

In a **live attenuated vaccine**, the virus is still viable though its virulence is greatly diminished. In contrast, in inactivated vaccines, virus particles are used.

A **hydrophobicity** plot is a quantitative analysis of the degree of a protein's attraction or repulsion to water. When analyzed, the plot gives partial information about the protein's structure.

Dengue virus ravages subtropical regions and is structurally similar to flavivirus with its lipid bilayer envelope and RNA genome. After its breakout in recent years, it has infected an estimated 390 million people per year.

Dengue infection stimulates the production of numerous low-affinity antibodies that can potentially worsen the infection.

Creating and filtering artificial dengue E and M protein sequences against multiple Dengue virus serotypes to develop an effective vaccine

Oliver Cai

Annotated by Wasi Shahriar, Chris Dong, Justin Leung, Md Shad, Nina Shin

Abstract

Like all viruses, **dengue has an incredible capacity for evolution**, and currently has four known genetically distinct **serotypes**. There is a great demand for an effective vaccine that can incite an immunological response to all dengue serotypes. I employed a **Hamming distance based genetic algorithm** to generate varied sequences for dengue E and preM surface glycoproteins, (both of which are involved in the virus's ability to infect cell membranes), in order to create an efficient **live attenuated vaccine** that has enough cross reactivity to defend the body from all dengue serotypes. **Hydrophobicity analysis** was employed in order to ensure that the artificially-generated proteins have a good chance to fold during an actual lab test. The centroid hamming distance successfully generated 50 sequences for E and preM proteins, each averaging a hamming distance of 123 and 33 respectively from the original parent sequences. The sequences demonstrated high phylogenetic correlation with existing dengue sequences from all serotypes. If this set of sequences is successful then we can assume that this computer generation procedure may be the next solution to fighting the world's evolving viruses.

Introduction

Dengue virus is a disease with symptoms that can range from mild headaches and fevers (Dengue fever) to more severe issues such as shock and internal fluid leakage. Dengue fever's distinguishing trait is its deadly secondary infection, and a confounding factor is that dengue virus comes in four major serotypes that are distinct enough so that no one specific antibody created by the immune system can neutralize all of them. Thus, people who contract a secondary infection are much more likely to have their symptoms develop into dengue shock syndrome or dengue hemorrhagic fever due to a combination of viral uptake and ineffective antibodies.

Fortunately, several different live attenuated vaccines(LAVs) are currently in development: DENVax uses the backbone of one serotype and replaces the essential membrane proteins with the proteins of other serotypes, effectively covering all four serotypes atonce(9). A diagram of the virus backbone can be seen in Fig1A(4). LAV Delta 30 uses all four serotypes, with 30 base deletions at the

once(9). A diagram of the virus backbone can be seen in Fig1A(4). LAV Delta 30 uses all four serotypes, with 30 base deletions at the end to minimize symptoms(10). A diagram of the backbone example is shown in Fig 1B (4). There are also recombinant DNA vaccines that specifically cause the body to produce dengue membrane proteins to provoke the production of antibodies(11). An example is shown in Fig 1C(4). Only one vaccine is currently FDA approved, the CYD tetravalent dengue vaccine which utilizes a **flavivirus** backbone and replaces several crucial proteins with dengue counterparts(4). A diagram can be seen in Fig 1D. In study, many of these vaccines elicit successful immune responses to dengue infections, however the antibody production and the quality of recovery has varied significantly between serotypes, indicating that better approaches can be made (see Fig 1E). My study approaches the development of a vaccine through artificially generated protein sequences. Proteins derived from a phylogenetic algorithm were created in order to cover a wide range of influenza clades. My study aimed to employ a hamming distance genetic algorithm (7) to create "centroid sequences". The program then tries to minimize this distance by generating numerous mutated sequences until the hamming distance hits a constant minimum (an equidistant difference from all parent sequences, therefore coining the term "centroid sequences"). To further increase the possibility of broad **cross-reactivity**, a program based on hydrophobicity(3) was also developed to further increase the chance of protein folding. The success of the results were measured by the creation of phylogenetic trees with the centroids in order to observe their successful genetic correlation with all four serotypes. These trees were created both with FASTTREE and Bayesian programs (5). Dengue proteins E and preM surface proteins were chosen because of their function as i**ntegral viral membrane proteins** as well as their similarity across all four serotypes.

A **flavivirus** is a family of RNA viruses that usually have arthropod vectors and cause serious human diseases, including dengue.

Cross-reactivity is when an observed agent causes a reaction that is different from the expected reaction. More specifically, it is when an antibody can recognize two antigens with similar structures. It measures the degree to which different antigens can cause an immune reaction.

Methods

Obtaining sequences

Sequences were obtained via genbank. A collection of dengue E and preM protein sequences from all four serotypes were downloaded in FASTA file format. After eliminating all duplicate sequences, there were a total of 82 E sequences and 19 preM sequences.

Integral Membrane Proteins are proteins that are permanently attached within a biological membrane of a cell.

Aligning and Parsing Sequences

In order to properly execute the hamming distance algorithm, the original sequences were parsed and aligned using **Biopython** and **MUSCLE**.

For most of the E and preM sequences, the length was about the same. For E, the protein length averaged 300 amino acids, and for preM, the protein length averaged 150 amino acids. The MUSCLE alignments produced FASTA files with few gaps, already indicating a high genetic correlation between protein types. ClustalW files and

Biopython is a collection of Python tools to be used in computational biology, bioinformatics, FASTA files, and MUSCLE alignments. FASTA files contain text representing nucleotides or amino acids.
MUSCLE is a tool used for sequence alignment of proteins and nucleotides.

Seqmagick is an bioinformatics formatting tool that converts a large amounts of script to biopython.

Phylogenetic trees are visual representations of evolutionary relationships between multiple species of organisms. Data used to sort these trees can range from physical to genetic/molecular information.

MrBayes is a program utilized for Bayesian inference, which is a method of statistical inference using Bayes' theorem. MrBayes is also employed for model choice throughout the range of phylogenetic and evolutionary models. MrBayes uses Markov chain Monte Carlo (MCMC) methods to estimate the posterior distribution (the conditional probability after accounting for relevant information).

FastTree is a way to store sequence profiles of internal nodes in phylogenetic trees. Using the profiles and heuristics, it applies "Neighbor-Joining" while determining candidate joins. It then uses interchanges between the nearest neighbors to decrease the length of the tree.

Bayesian trees are the models created using MrBayes for data prediction and exploratory modeling. The Bayesian trees are very powerful instruments within the various data mining methods, allowing predictions of different types of outcome that range from quantitative and qualitative data to the time until an event occurs.

Immutable regions are locations in the amino acid sequences of proteins that are constant and shared among a collection of proteins.

The **hydrophobicity analyzer** is a python terminal line command program based on the kyte-doolittle hydropathy scale.

Hydropathy correlations reveal information about the structures and functions of different proteins by examining the anticodons and the amino acid excess amounts

Nexus files were also produced through **seqmagick**. These files were later used in the making of **phylogenetic trees** through **MrBayes** and **Fasttree**.

Creating Trees

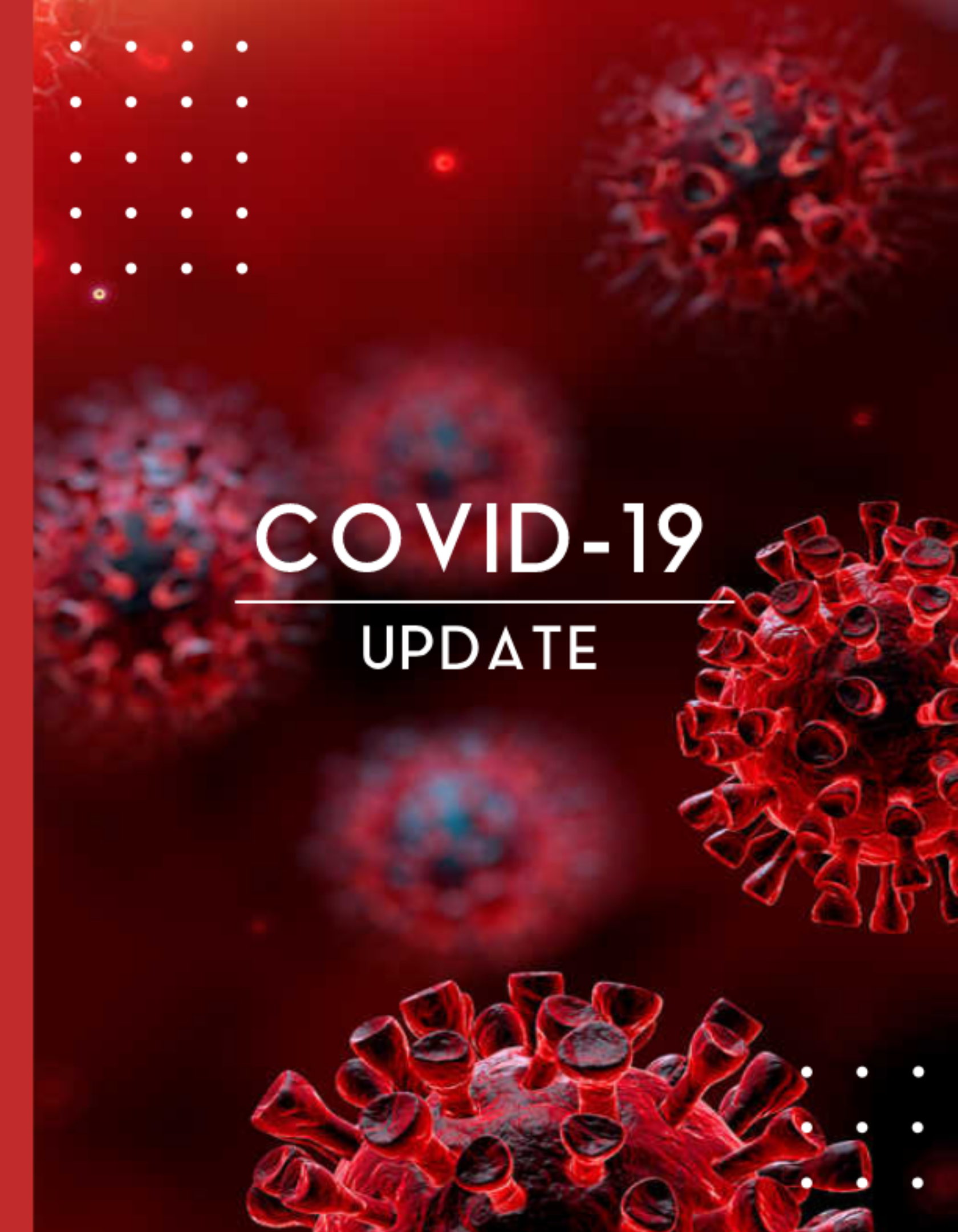
Trees were created to get a baseline observation of phylogenetic correlation. Fasttree produced a tree that portrayed very high genetic correlation, with each clade being grouped with its respective serotype sequences. With Fasttree, the clustalw files were used.
To further refine the accuracy of the trees, MrBayes was used to generate **Bayesian trees** (5)(6). Since these trees were produced through an entirely independent manner with the Monte Carlo Markov Chain(MCMC), they would allow us to determine the most accurate support values for the tree branches. In this method, nexus files were used.

Creating Artificial Sequences/Centroids

Artificial sequences were generated with an evolution algorithm designed to minimize hamming distance. Sequence hamming distance is calculated by summing the number times an amino acid diverges from the parent sequence. Starting with a set of real sequences, the program “remembers” **immutable regions**, thus designating regions where mutations can occur without disrupting protein function. With each generation, a new "mutation" takes place. After several generations, the lowest hamming distance measurement will eventually become constant, indicating that a new mutation will not be able to lower the hamming distance any further. This lowest hamming distance sequence is then outputted as a “centroid sequence”.

Refining results with Hydrophobicity Analyzer:

Centroid sequences and original sequences were inputted, analyzed, and filtered in the **hydrophobicity analyzer**. The 10 centroid sequences with higher **hydropathy correlations** were outputted in a FASTA file. Source code can be found here https://github.com/OliverCai0/Hydrophobicity_Correlation. The program also contains additional modifiers so that the user may adjust settings to tweak things such as error range, overlapping positions, number of filtered sequences, etc.


COVID-19
UPDATE

COVID-19 Update

Background

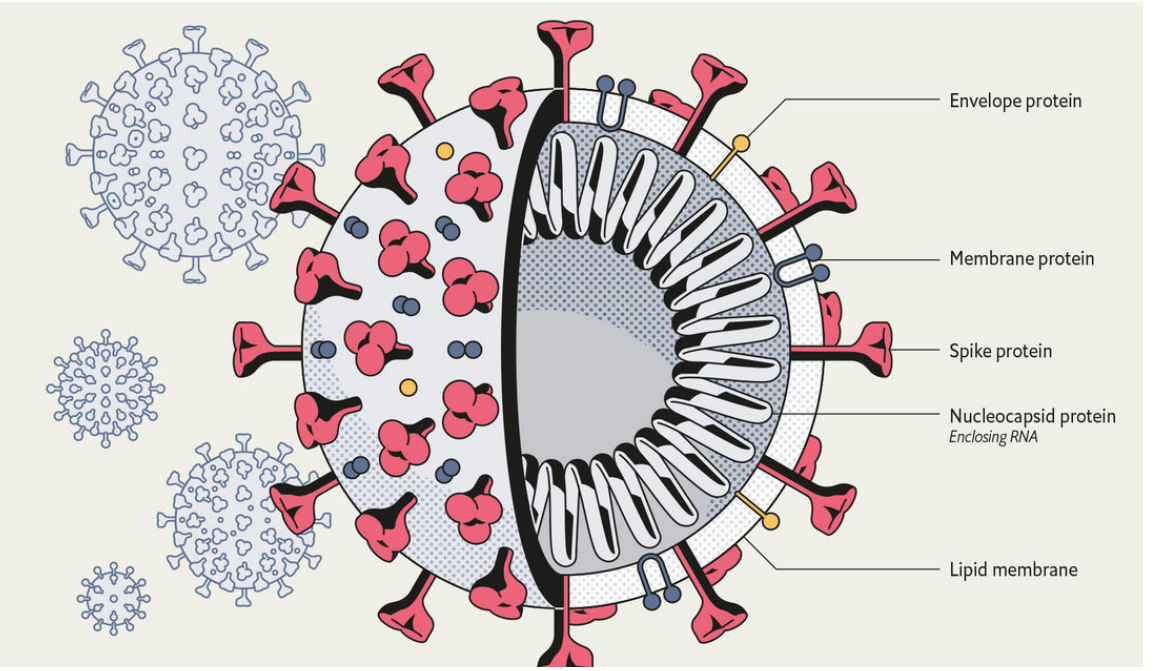
Unprecedented is what comes to mind when discussing the global COVID-19 pandemic. Over a span of a mere four months, the COVID-19 pandemic has spread to over 185 countries, shut down many countries, schools, and workplaces, and irreversibly changed the trajectory of global progression.

COVID-19 is a respiratory illness caused by the novel coronavirus SARS-CoV-2, a strain of severe acute respiratory syndrome-related coronavirus. Coronaviruses (are a large family of viruses, causing diseases ranging from the common cold to more severe illnesses. In the past, occurrences of SARS and MERS—each under the coronavirus family—have spread locally, with contained and short outbursts. The novel coronavirus, however, is a new strain that previously has not been identified in humans. The virus has zoonotic origins with close genetic similarity to that of bat coronaviruses.

To infect a human host, the virus gains entry into the human's cells through openings in the body through crown-like projections from its surface. The virus then uses the machinery of infected cells to produce copies of themselves, replicating and spreading to new cells. Once the virus enters the body, it attaches to the ACE2 receptor and fuses its membrane with the infected cell, incorporating its RNA. The infected cell then makes new copies of the virus that are expended through the body, although the infected individual may be asymptomatic for up to two weeks upon contracting the virus.

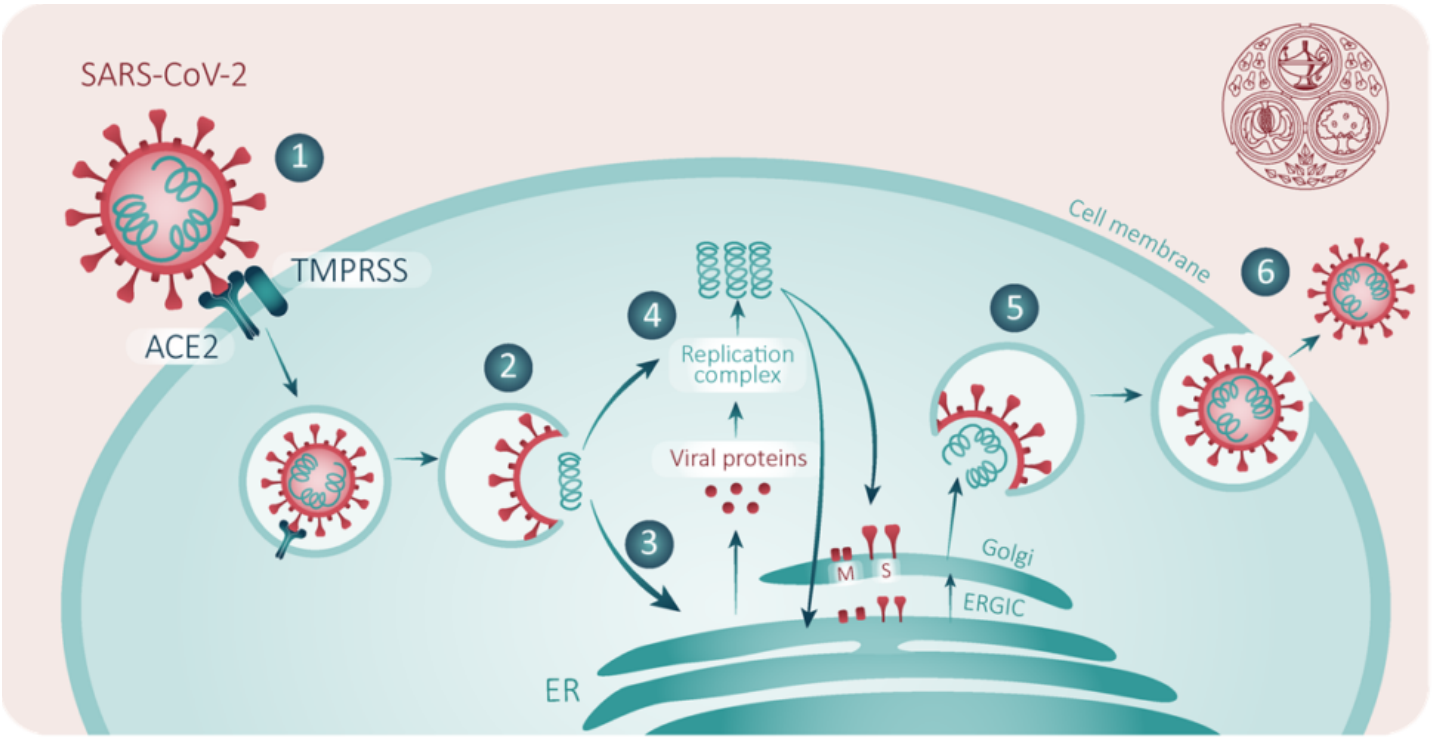
At this stage, typical symptoms range from fever, coughing, sore throat, weakness, fatigue, muscular pain, or appetite loss. The progression of the disease, however, is typically due to an overreaction of the immune system. As the virus enters lung cells, the immune system response can destroy lung tissue and cause major inflammation, leading to pneumonia. In the most critical cases, the body starts to fail.

The body experiences severe respiratory problems, organ failure and septic shock; all which can lead to death. As COVID-19 continues to spread, governments and labs around the globe have responded accordingly. The dedicated pursuit of effective treatments, vaccines, and inventions have ensued as a result of the pandemic, making the passing of COVID-19 a closer reality with each passing day.



The structure of the COVID-19 virus

Sources: Manuel Bortoletti et ad 'The Economist' 2020



1 Spike protein on the virion bind to ACE2, a cell-surface protein. TMPRSS2, an enzyme helps the virion enter 2 The virion relases its RNA 3 Some RNA is translated into proteins by the cell's machinery 4 Some of these proteins form a replication complex to make more RNA 5 Proteins and RNA are assembled into a new virion in the Golgi nand 6 a virion is released.

Sources: Song et al., 'Viruses', 2019; Jiang et al., 'Emerging Microbes and Infections, 2012

Vaccines

In the heat of the SARS-CoV-2 outbreak, the efforts to slow the spread of the virus as well as building an immunity in the general population has greatly increased. Despite the intensity of the outbreak, vaccines, such as antibody tests or spike protein tests, prove to be a possible method of immunity.

The University of Pittsburgh School of Medicine has taken a different approach on helping build population immunity to this pandemic: the MERS-CoV S1 clinical trial. This particular vaccine targets the strain of coronavirus that causes the Middle Eastern Respiratory Syndrome. Unique to itself, the method utilizes a microneedle patch that injects viral spike proteins into the bloodstream, causing a response built off of adaptive immunity. Similar to other pathogenic infections, the spike protein of a pathogen is used as a catalyst for the activation of B cells immunoglobulins that work to produce more memory cells and plasma cells. So far, animal testing has been implemented to test the effectiveness of this vaccine. Mice that were exposed to the MERS-CoV and injected with the same spike proteins had significant efforts of viral neutralizing activity, such as the production of antibodies.

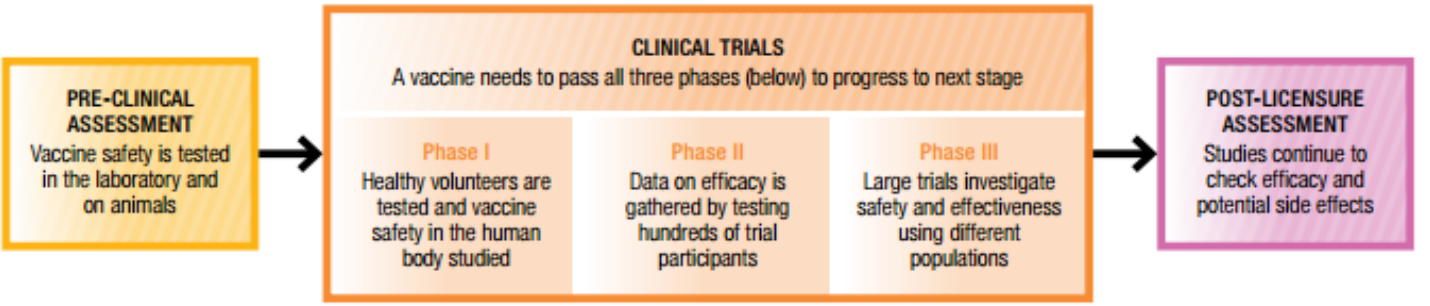
However, significant changes in the antibody levels were found between the 23rd and 55th week after the initial exposure to the spike protein. Although it is possible that this time could be shortened or elongated in humans, the University of Pittsburgh School of Medicine has not come to a conclusion about this specific time period at the moment.

This vaccine, SARS-CoV-2 mRNA-1273, takes a different approach. By utilizing the patient's cells' protein production machinery, mRNA, the vaccine stimulates an immune response through the production of spike proteins. mRNA vaccines are particularly useful since some viruses do not possess a lysogenic cycle. While a virus with a lysogenic cycle, it is able to insert its own genome into the host cells, making it difficult for the cells to perform normal cell function. This is common for diseases such as HIV. Once the mRNA that transcribes the spike protein is injected into the body, the somatic cells' ribosomes will produce the viral spike proteins. Once the proteins are released into the extracellular environment and activate the proper plasma B cells that produce antibodies. The antibodies are able to bind to the spike proteins on an active virus, given that the patient becomes infected.

One advantage is that the mRNA degrades overtime, causing the immune response to halt, but just for memory cells to be produced in case of a second exposure. The mRNA will induce spike protein production for a small period of time, just enough time to produce memory cells and plasma cells.

Building off of other vaccines developed for viral outbreaks such as the Ebola outbreak, the adenovirus method has also been used in multiple trials as an attempt to combat SARS-CoV-2. CanSino Biologics and the Academy of Military Medical Sciences, fostering the trials of Ad5-nCoV, launched their clinical trials on March 17th. It involves a series of healthy adults that have no underlying medical conditions prior to the start of the trial. 500 subjects were used for this clinical trial, with 250 of them receiving a middle dosage of the vaccine, 125 in a low dosage group, and 125 in a placebo group. Proteins that will stimulate an B cell immune response by binding to surface immunoglobulins. The goal of this trial is to determine the effectiveness of the vaccine in respect to the amount that is injected into the body. Adenovirus type 5 has been used in other viruses, which served as gene delivery vectors. Seeing as this indicates the ability of adenovirus type 5 to be easily manipulated genomically, scientists have altered the genome of the virus to cause it to express COVID-19 spike proteins that will stimulate a B cell immune response by binding to surface immunoglobins.

As the COVID-19 pandemic starts to deescalate through social distancing efforts and stay at home orders from the government, many laboratory researchers are discovering newly invented vaccines that can potentially prevent the pandemic from reoccurring later on.



The process of mandating a vaccine by the FDA in America
Sources: Australian Academy of Science: The science of immunization

Inventions

In light of the ongoing COVID-19 pandemic, innovators around the world are working to minimize the spread of the virus, both by aiding health care workers and protecting those unaffected by the virus. Due to the highly infectious nature of the virus, the task of slowing its spread is crucial to overcoming the pandemic.

Winsun 3D is a Shanghai-based company that specializes in printing 3D architecture. The company developed a 3D printed isolation ward to aid in hospital overpopulation and isolation efforts. Each 10-sq meter isolation pod can be printed in the span of just two hours. Buildings like these can be printed quickly and flexibly moved since no foundation is needed beforehand. The company utilizes non-traditional, eco-friendly resources, and equipping each ward with an ecological toilet broken down by organic bacteria. These isolation wards accommodate quarantined individuals and medical staff, aiding in reducing hospital overflow.

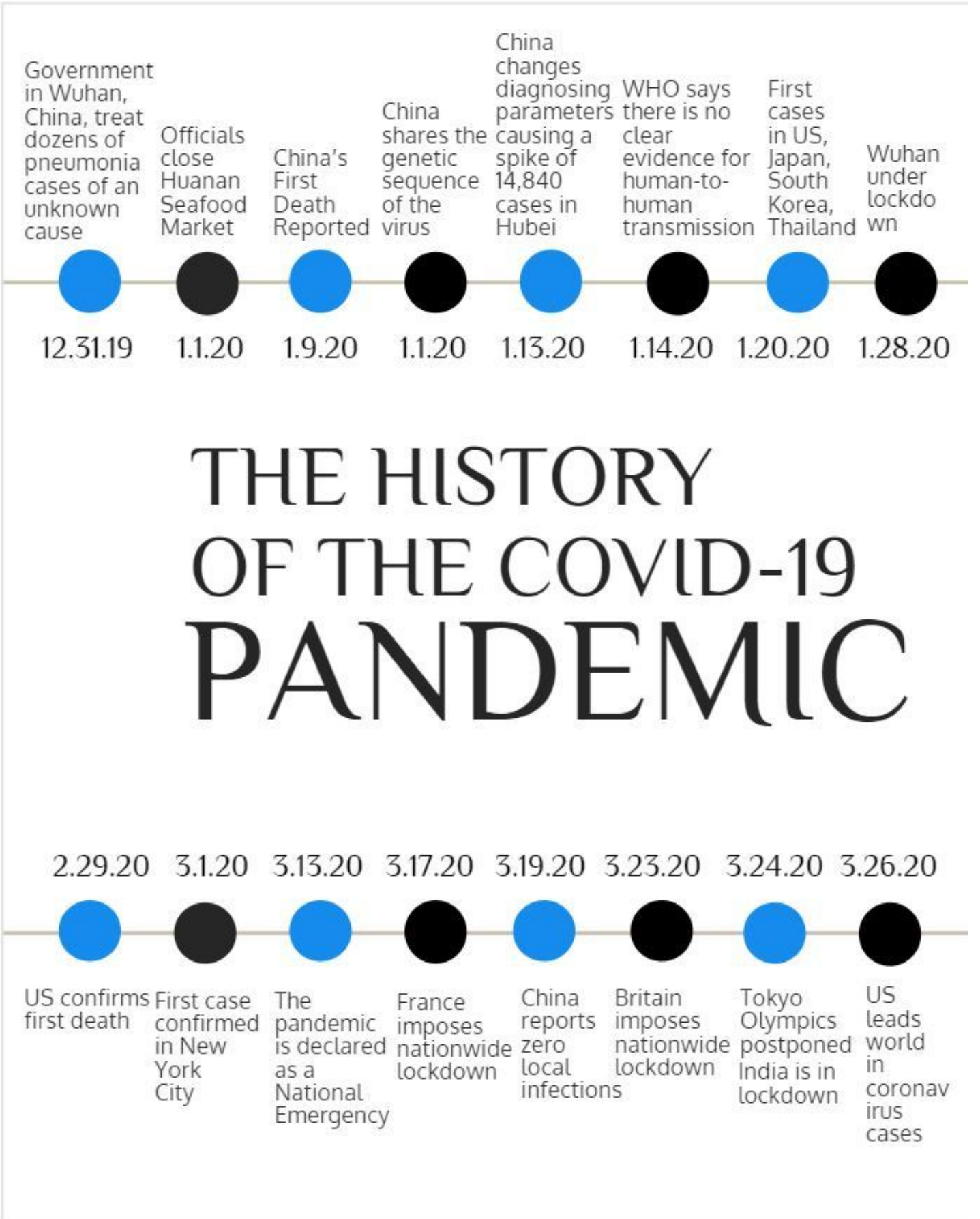
These specialized robots reduce transmission rates before infections occur in the first place. Developed by UVD Robots, a Danish company, their robots are capable of utilizing powerful, short wavelengths of ultraviolet-C (UVC) to sterilize hospital wards. The robot consists of a base equipped with Light Detection and Ranging (LiDAR) senses and an array of UV lamps and cameras. By using LiDAR the robot is able to move through a series of pre-mapped locations in a room. The remote sensing method creates a 3D elevation map of a particular surface by using light in the form of a pulsed laser to measure relative distances. Afterwards, the robot relies on Simultaneous Localization and Mapping (SLAM) to navigate and operate autonomously.

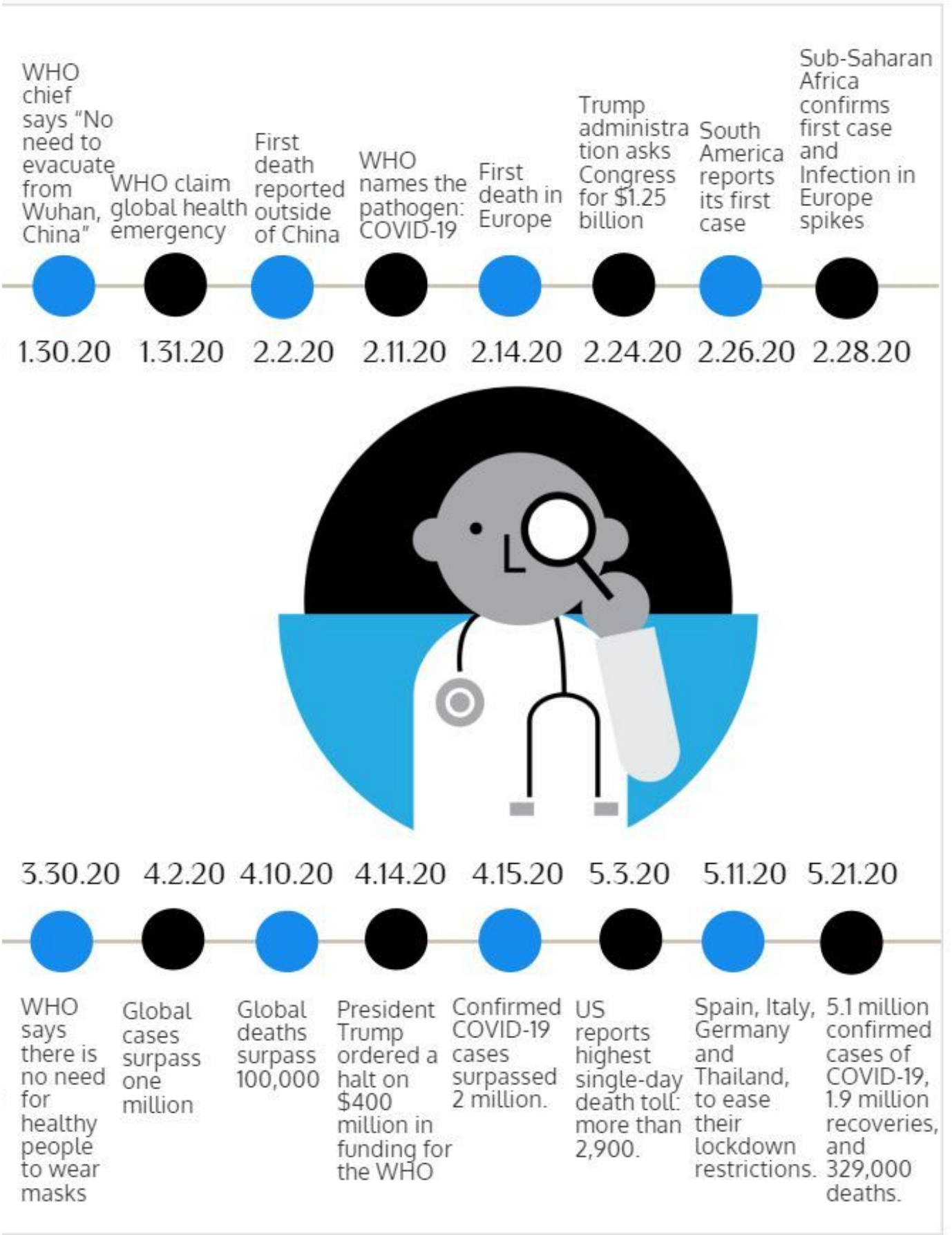
Workers installing a Winsun 3D ward
Source: UV Disinfection With Pulsed Xenon

Similarly, the US-based company Xenex creates LightStrike robots that use pulsed xenon ultraviolet (PX-UX) as an alternative to traditional UV methods. PDX-UV is typically emitted in short, high-intensity pulses and possesses a high efficacy compared to other forms of UV (including mercury UV) due to its greater intensity and broader spectrum. With greater germicidal UV intensity, the LightStrike robots can effectively disinfect a patient room in less than 20 minutes.

Contact tracing is another process designed to help prevent the spread of a virus by identifying affected individuals and subsequently notifying those in close proximity. Many mobile apps have been developed in various nations with massive rates of usage. The Corona 100m (Co100) App was developed in South Korea and launched on February 11th, 2020. Co100 alerts its users when in vicinity of a 100-meter radius of an infected patient. Since its launch, the app has been downloaded over 1 million times. Co100 collects data from public government databases to show the date a patient was diagnosed with COVID-19, including the patient's age, gender, nationality, and locations visited.

Yet as coronavirus surveillance escalates, issues of personal privacy and security surface. On April 10th, Apple and Google announced a contact-tracing system utilizing Bluetooth Low Energy (Bluetooth LE) radio technology in both iOS and Android smartphone operating systems. Using Bluetooth, smartphones are notified of affected areas and individuals nearby. The new system would establish a voluntary contact-tracing network while retaining anonymity in physical location. The functionality of these contact-tracing methods, however, rely on people's own incentives to report disease and conditions, respond to warnings, and mitigate the spread of the virus to fully establish.


Government in Wuhan, China, treat dozens of pneumonia cases of an unknown cause
12.31.19
Officials close Huanan Seafood Market
1.1.20
China's First Death Reported
1.9.20
China shares the genetic sequence of the virus
1.1.20
China changes diagnosing parameters causing a spike of 14,840 cases in Hubei
1.13.20
WHO says there is no clear evidence for human-to-human transmission
1.14.20
First cases in US, Japan, South Korea, Thailand
1.20.20
Wuhan under lockdown
1.28.20
THE HISTORY OF THE COVID-19 PANDEMIC
US confirms first death
2.29.20
First case confirmed in New York City
3.1.20
The pandemic is declared as a National Emergency
3.13.20
France imposes nationwide lockdown
3.17.20
China reports zero local infections
3.19.20
Britain imposes nationwide lockdown
3.23.20
Tokyo Olympics postponed India is in lockdown
3.24.20
US leads world in coronavirus cases
3.26.20


WHO chief says "No need to evacuate from Wuhan, China"
1.30.20
WHO claim global health emergency
1.31.20
First death reported outside of China
2.2.20
WHO names the pathogen: COVID-19
2.11.20
First death in Europe
2.14.20
Trump administration asks Congress for $1.25 billion
2.24.20
South America reports its first case
2.26.20
Sub-Saharan Africa confirms first case and Infection in Europe spikes
2.28.20
3.30.20
WHO says there is no need for healthy people to wear masks
4.2.20
Global cases surpass one million
4.10.20
Global deaths surpass 100,000
4.14.20
President Trump ordered a halt on $400 million in funding for the WHO
4.15.20
Confirmed COVID-19 cases surpassed 2 million.
5.3.20
US reports highest single-day death toll: more than 2,900.
5.11.20
Spain, Italy, Germany and Thailand, to ease their lockdown restrictions.
5.21.20
5.1 million confirmed cases of COVID-19, 1.9 million recoveries, and 329,000 deaths.

2019

2020

STUYVESANT
RESEARCH CLUB

345 Chambers Street
New York, NY 10282
stuyresearch.weebly.com