Fall 2017
Vol. 10, Issue 2

thurj

the harvard undergraduate research journal

MODULATING MYC ACTIVITY

a comparative study across cell lines of novel small molecule inhibitors of MYC transcription, which is overexpressed in human cancers

evaluating Brazil's affirmative action policy

the biological aspects of autism spectrum disorders

the fiscal impact of immigrant characteristics

Congratulations to

Muhammed Ors

for winning

Best Manuscript

Muhammed Ors is a Harvard graduate of the Class of 2017. Muhammed concentrated in Molecular and Cellular Biology with a secondary in Islamic Studies and conducted research in the Koehler Lab at the Koch Institute of MIT. Outside of academics, Muhammed was President of the Voice Actors Guild and Co-chair of the Pforzheimer House Committee. A second-generation Turkish American originally from New Jersey, Muhammed currently works in Boston as a Research Technician in the Livingston Lab at Dana Farber and is performing some of the first ever mouse studies done on the BRCA1 isoform known as BRCA1-IRIS.

December 2017

Dear Harvard Community,

We are delighted to present the Fall 2017 issue of The Harvard Undergraduate Research Journal (THURJ), a student-run biannual publication dedicated to showcasing outstanding research from the Harvard undergraduate community. For the past nine years, we have been proud to publish high-quality, original research from a wide range of disciplines. In this issue, our authors explore topics as far-ranging as the biological aspects of autism spectrum disorder and the fiscal impacts of immigration. This year's Best Manuscript award goes to Muhammed Ors '17 for his outstanding manuscript on small molecule modulators of MYC in cancer cell lines.

As the journal grows and evolves, we are working to increase THURJ's visibility in the Harvard community and strengthen our reputation as a professional-quality journal. We recently introduced a new online submission platform designed to streamline the peer review process and better protect our student researchers' intellectual property. This semester we also expanded our faculty review process from within Harvard to all US universities. This new policy will allow us to find the very best faculty reviewers for all of our submissions and ensure that we are publishing top-notch research. THURJ continues to work with organizations on campus that share our vision, as well as undergraduate research journals at our peer institutions. Every semester, we strive to reach out to the larger scientific community as a platform for Harvard undergraduates to share their discoveries with the world.

As always, this work would not be possible without the incredible insight, dedication, and support of our THURJ team and faculty advisers. We want to thank the student and faculty reviewers, staff writers, and designers who put so much effort into creating, editing, and polishing this issue. In particular, we would like to thank our faculty advisor Professor Guido Guidotti for his guidance and wise counsel. We are also tremendously grateful for continued support from FAS Dean Michael Smith, FAS Dean of Science Jeremy Bloxham, FAS Dean of Social Sciences Claudine Gay, Harvard College Dean Rakesh Khurana, Professor Richard Losick, Provost Alan Garber, Vice Provost for Research Richard McCullough, Professor Steven Freedman, HMS Dean George Daley, Harvard Catalyst, the Office of the President, Harvard SEAS and Harvard College.

We are delighted to present our newest issue and share this outstanding research with the Harvard community. Enjoy!

Sincerely,

Katie Kixmoeller
Co-Editor-in-Chief

Bryan Peacker
Co-Editor-in-Chief

Contents

Research

Features

Contact

General: contact@thurj.org
Advertising: advertising@thurj.org
Subscribing: subscriptions@thurj.org
Submissions: submissions@thurj.org
Website: http://www.thurj.org

About Us

The Harvard Undergraduate Research Journal (THURJ) showcases peer-reviewed undergraduate student research from all academic disciplines. As a biannual publication, THURJ familiarizes students with the process of manuscript submission and evaluation. Moreover, it provides a comprehensive forum for discourse on the cutting-edge research that impacts our world today.

At its core, THURJ allows students to gain insight into the peer review process, which is central to modern scientific inquiry. All THURJ manuscripts are rigorously reviewed by the Peer Review Board (consisting of Harvard undergraduates), and the top manuscripts that they select are further reviewed by Harvard graduate students, post-doctoral fellows, and professors. This process not only stimulates faculty-student collaboration and provides students with valuable feedback on their research, but also promotes collaboration between the College and Harvard's many graduate and professional schools. In addition to publishing original student research papers, THURJ is also an important medium for keeping the Harvard community updated on research-related news and developments.

About the Cover

This cover features an image of a cancer cell. It represents our Best Manuscript for Fall 2017, "Pursuing Synergy: Combining a Pan-Class PI3K Inhibitor with Novel Small Molecule Inhibitors of MYC." Author Mohammed Ors wrote a paper on his research in Koehler Lab, where he used novel small molecule modulators of MYC activity in a comparative study across cancer cell lines against inhibitors of MYC. Overexpression of MYC is one of the most common mutations in human cancers.

Source for summaries images: adobestock.com. Source for the cover photo: istockphoto.com.

Evaluating Brazil's Affirmative Action Policy

by Ike Okonkwo p.11

Over the past few decades, higher education in Brazil has become a trending topic of concern as each year Brazil's competitive public universities grow in applicants yet remain lacking in diversity initiatives. In 2012, Brazil passed a federal affirmative action law which reserved spots in these institutions for both low-income and minority students. This education policy set a legal precedent for Brazilian's to identify themselves racially, sparking a national debate around race and inequality in the education system. This study was designed to access the beliefs and attitudes of Brazilian students about their experience in public universities and aimed to evaluate how the 2012 quota law was received by them. According to the qualitative data and narratives of the students in interviews, we confirm evidence of an internalized discourse on race and socioeconomic status that dictates how these factors interact with their access to higher education and their course selection.

Pursuing Synergy: Combining a Pan-Class PI3K Inhibitor with Novel Small Molecule Inhibitors of MYC

by Muhammed Ors p.16

The overexpression of the transcription factor MYC is one of the most common mutations in human cancer. Recently, a study on the oncogene PI3K, a kinase, showed that MYC drove resistance to inhibition of the oncogene by the inhibitor compound GDC-0941. The study showed that this resistance was preventable by using the compound JQ1 to indirectly suppress MYC activity by inhibiting another protein called BRD4. Here we use novel small molecule modulators of MYC activity developed by the Koehler Lab in a comparative study across human breast and brain cancer cell lines against the aforementioned inhibitors GDC-0941 and JQ1. The hypothesis of this study was that the small molecule modulators would be more effective than the other compounds in reducing cancer cell viability due to being able to directly inhibit MYC activity. Ultimately, cell viability assays in the breast lines confirmed that the novel small molecule modulators of MYC reduces the viability of these cell lines, but not yet to a degree significantly greater than the other compounds tested. Furthermore, none of the compounds seemed to be effective in reducing viability of the brain line.

Understanding Immigration: A Closer Look at the Fiscal Impact of Immigrant Characteristics

by Jean Thirouin p.28

This paper attempts to further research on the fiscal impact of immigration. My findings are illustrated by computing the net fiscal impact, in present value terms, of admitting one additional immigrant, conditional on education, gender, and age at the time of immigration. I demonstrate that the average immigrant arriving past age 34 has a lifetime negative fiscal impact. Additionally, a college educated immigrant arriving past age 52 will have a lifetime negative fiscal impact while a non-college educated immigrant will roughly have a lifetime negative fiscal impact, regardless of age at arrival. Further, I confirm that age at arrival matters, and determine that arrival prior to working age influences educational attainment. Finally, I provide a household life-cycle model that sheds light on the fiscal contribution of immigrating families. All immigration referred to in this paper is legal.

A Review of the Biological Aspects of Autism Spectrum Disorder

by Jennifer Bi, Puneet Gupta, Katherine Miclau, Sarah Ryan, Yong Shen, Katherine Venturo-Conerly, Akash Wasil p.37

Autism Spectrum Disorder (ASD) is a relatively common developmental disorder, resulting in social, behavioral, and cognitive differences. ASD presents as many different clusters of symptoms of varying severity. It is difficult for physicians to diagnose Autism Spectrum Disorder (ASD) in children due to the many forms in which it can manifest. Factors that affect how ASD presents include the patient's cultural upbringing, socioeconomic class, and co-morbid disorders. The influence of cultural, social and environmental factors are all potential causes of misdiagnosis, and all factors that physicians and researchers must take into account when interacting with these individuals. In this literature review, we explore many common sociocultural causes of ASD delayed diagnosis and misdiagnosis. We argue for more nuanced diagnostic approaches, which will allow for earlier and more accurate diagnoses. Early and accurate diagnosis of ASD allows for early treatment, which is essential for the healthy development of children with ASD.

Research

Evaluating Brazil's Federal Affirmative Action Policy

Ike Okonkwo '18

In 2012, Brazil passed Lei de Cota — a federal law which implemented a U.S. style of affirmative action in its public universities. This policy reserved seats for underrepresented minority and low-income students and established a legal precedent for Brazilians to racially identify. In this text, we examine the effects of Lei de Cota on Brazilian students through a microscopic lens. Using both qualitative and quantitative methods, we engage university students to learn about their experience with quotas on campus. Through student responses to questionnaires and interviews, we reveal an internalized racial & socio-economic discourse as it relates to access to higher education education and even course selection.

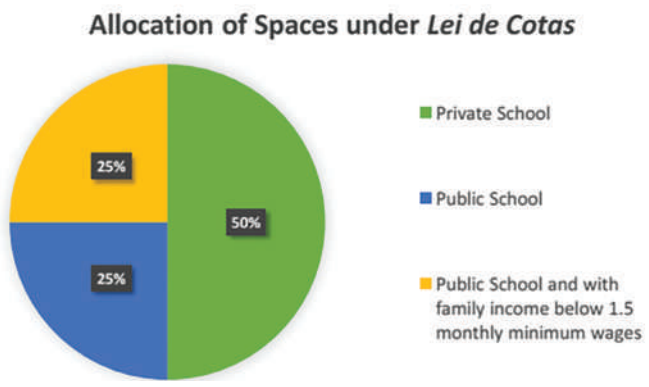
Expanding access to Brazil's institutions of higher education, particularly for historically underrepresented racial minorities and low-income students, has long been a topic of interest in Brazilian affairs. Although racial mixing in Brazil certainly produced a spectrum of races and phenotypes that span economic classes, racism and colorism are still present in society. It can be argued that the racial democracy in Brazil asserts the absence of racism as a byproduct of the country's history with racial mixing[1]. However, evidence of racial inequality specifically in the Brazilian education sector can easily be found. In Chapter 1 of Maria Susana Arrosa Soares', *Higher Education in Brazil*, Soares portrays the foundation of Brazilian colleges and their backwardness. When Brazil became independent in 1822, education was restricted to the descendants of Portuguese families. In 1931, President Getúlio Vargas created the Ministry of Education opening courses that perpetuated this exclusion in the education sector[2]. Even in 1934, with the foundation of the Federal University of São Paulo (USP, widely regarded as the highest quality university in Brazil), the 'Paulista' elite continued having privileged access to professional courses in medicine and law[3]. Clearly, this complex colonial history influenced a gap in who had political power, financial stability, and job opportunities. This raises the questions: How, if at all, can we repair these inequities? More, can this be accomplished through Affirmative Action (AA) policy?

In order to make sense of Brazil's AA policy, it is first necessary to understand Brazil's various existing models of higher education. In this text we will focus on Brazil's public universities, which must adhere to Brazilian federal AA law. By examining the work of Brazilian scholars like Maria Soares, we can conceptualize the various models of financing for these institutions and the implications that they pose for under-represented students. These institutions include both state and federal colleges, referred to as public universities. For state colleges, the state is the principal source of funding for the institution, while the federal colleges are maintained by the federal government and depend on its tax revenue as the primary source of income[4]. Since education in these public institutions is free to its students, the application process to attend them remains supremely competitive. This is not to say, however, that Brazilian private colleges are necessarily uncompetitive. In fact, Soares highlights how private universities are also growing uncontrollably each year, citing the 1,004 private institutions within Brazil serving 1.8 million undergraduates (roughly 2/3 of the country's student population in higher education)[5]. Still, given that Brazilian private institutions do not offer free tuition, they are less preferred or financially inaccessible to many students, especially those from low-income households.

Current trends in Brazilian education demonstrate an increase in university applicants accompanied by increasing debate over quota policies. It is worth recognizing the increased discourse surrounding AA policy as it has turned into a highly controversial topic in Brazil. Studying the student population and implementation of AA at federal institutions helps us better understand the competitive nature of Brazil's school system as it relates to socio-economically disadvantaged students. Brazil's Census of Higher Education confirms a recent growth in the enrollments to its colleges stating: "the percentage of people attending university is almost 30% of the Brazilian population in the age group of 18 to 24 years, and 15% is at the age theoretically adequate to attend this level of education."[6] This information (perhaps) reflects a general distrust in the labor market such that more people feel that higher education is needed to successfully secure employment. Even distance learning courses have been increasing in annual enrollment[7]. The creation of more vacancies on college campuses generates expenditures that are not sustainable in the long-term, exacerbating the need to create effective access to education and to rethink plans to increase the quality of all the available institutions (both private and public). To that end, I propose to approach the topic of affirmative action in Brazil taking into account its racial and economic quotas and the ways in which they operate.

Study

It is important to review the implementation of Brazil's affirmative action policies in order to understand the demographic makeup of Brazil's public colleges. In 2012, Brazil's Federal Affirmative Action Law (Lei de Cotas) established a reserve of 50% of vacancies in federal institutions for students who had completed public high schools. Of the half of the reserved slots, 50% were earmarked for students with per capita family income of up to a minimum wage and a half[8]. That is to say, a total of 25% of the university spaces were reserved for students from low-income families. In addition, the law also implemented a racial distribution across the reserved spots among Black, Mixed-Race (pardos), and Native-Indian students with respect to the proportion of ethnic groups in each state. The adoption of racial sub- quotas was the most polemical of the policy as it established an unprecedented legal process for Brazilians to identify themselves racially[9]. This polarizing policy gave rise to two factions: the supporters of racial quotas and its opponents.



This infographic illuminates the distribution of spaces in Public Universities[10]

By the same token, we must stop to address the consequences of applying to or not applying to the universities' quota pools. For students who have entered university through the quota system, they risk being identified as 'quota holders' (cotistas). In recent years, quota holders have faced social challenges as negative attitudes about their status have manifested in forms of bullying including cases of physical violence used against them[11]. That is to say, being a quota recipient (likely) equates to having greater chances of being beaten, isolated, or having one's academic merits and/or ethnic identity questioned. These factors present a negative cost for minority students, discouraging their use of the quota policy. Conversely, for students who choose not to apply to the quota system, they run the risk of being less likely to receive a seat in a public university. All of these challenges to (AA) policy are significant because they limit the access to higher education, a potentially powerful agent in combating cycles of intergenerational poverty and increasing an individual's social mobility.

In short, when doing an analysis of the quota system in Brazil it is crucial to take all these details into account alongside the students' experiences. Therefore, we investigate the effects of the 2012 AA law. Specifically we ask: how has federal AA policy in Brazil been received by its students and how has it influenced students on a day-to-day basis? We then address which measures can be applied to improve its implementation. The data collected from this study suggests an internalized discourse regarding race and social status that reflects the need to change federal quota policies and potentially the funding of higher education in Brazil as a whole.

Methodology

Methods

This project used both quantitative and qualitative methods to explore the relationships between participants' socioeconomic profile and their daily experiences as students of public universities. For the quantitative portion, we used SPSS to breakdown correlations between student attributes like race, family income, quota status, college, and rating of the quota system. We utilize regression analysis on these student attributes and their personal ratings on a scale of 1-5 of the quota policy (from least to most favorable), attempting to explain the variance in the students ratings. It is important to note that our quantitative methods report on a small number of participants (50 students), which could potentially invalidate the nature of such methods.

However, we still find some statistically significant data that could at least justify the continuation of further studies of a larger magnitude. For the qualitative portion of our study, students wrote open-ended responses about their most present challenges (specifically on college campuses) and also had the option to participate in oral-interviews.

Participants

This study was conducted with the help of 50 participants who were randomly selected from various Brazilian universities including the Federal University of Amazonas, the State University of Amazonas, the University of São Paulo, the Federal University of São Paulo, and the Federal University of Pernambuco. Participants were recruited using fliers that were posted on their respective campuses. Of the 50 participants who responded to these postings, 41 completed the questionnaire and agreed to have their answers published. Figures (1-4) depict demographic details of these participants (see Appendix).

Instruments

The data for this study were collected through two research instruments: audio recordings of in-person interviews and an online questionnaire developed to access the students' opinions of the quota policy.

Quantitative Results and Discussion

The following tables (1-2) depict the most relevant data on the relationship between students' course selection and his/her attributes such as: their rating of the quota policy, level of family income, and status as a quota holder. What we can see from this information is that there is a negative and significant correlation between the courses of business administration and the ratings of the quotas. In other words, as the level of satisfaction with the quota policy increases, the number of administration students drops. Most of the administration students in this sample come from the University of São Paulo (USP) and are also responsible for the positive correlation between high-family income and enrollment in the administration course. That is to say, that with an increase in the number of USP administration students, there is an accompanying increase in family income and lower satisfaction with quotas. It is also interesting

Correlations

		Administration	Law	International Relations	Social Service	Tourism
Quota Rating	r	-.318*	-.128	.152	.269	-.128
	p	.049	.439	.356	.098	.439
	N	39	39	39	39	39
Family Income	r	.397*	.309	.136	-.334*	-.220
	p	.012	.056	.409	.038	.177
	N	39	39	39	39	39
Quota Holding Status	r	.387*	.115	.038	-.380*	.115
	p	.015	.487	.816	.017	.487
	N	39	39	39	39	39

Table 1: Relationship between student course selection and the students' attributes (ratings of the quota system, family income, and status as shareholder). *Correlation is significant at the 0.05 level (2-tailed).
Note: the attributive quota student was coded using 0 = Quota Holder and 1 = Not Quota Student

Correlations

		Federal University of Pernambuco	Federal University of São Paulo	Federal University of Amazonas	Federal University of São Paulo	State University of Amazonas
Quota Rating	r	.192	.207	-.227	-.318*	-.183
	p	.243	.207	.164	.049	.265
	N	39	39	39	39	39
Family Income	r	-.271	.281	-.157	.397*	-.316*
	p	.095	.083	.340	.012	.050
	N	39	39	39	39	39
Quota Holding Status	r	-.387	.076	.000	.387	-.082
	p	.015	.647	1.000	.015	.619
	N	39	39	39	39	39

Table 2: Relationship between university and the students' attributes (ratings of the quota system, family income, and status as shareholder). *Correlation is significant at the 0.05 level (2-tailed).

to note that a parallel trend exists with the law course, though it is not statistically significant (perhaps) due to the sample's fewer number of law students. Our sample of quota holders relative to other students is relatively small, rightfully so, as the percentage of total quota holders overall is much smaller in many courses. In short, these correlational data may reflect a negative attitude among students in the most competitive courses (administration and law) in relation to quotas. Conversely, we find that the less competitive course, social service, possesses the opposite relationship regarding student enrollment and family income; social service also possesses the largest number of quota holders in total.

The continuation of these correlations can be seen in table two. However, table 2 depicts the relationship between university and the students' attributes (ratings of the quota system, family income, and status as shareholder). The most relevant data here show a negative and statistically significant correlation between quota satisfaction and enrollment at USP. Again, this factor is largely due to the administration students in the sample. More, we see a positive correlation between students attending this university and higher family income. On the other hand, we find a negative correlation between the enrollment at the University of the State of Amazonas and wealth. All of this is to say that our data confirms a relationship between the level of resources students have, what the student chooses to study, and where they choose to study.

Finally, we analyze the multi-variable regression that uses students' ratings of the quota system as a dependent variable and their attributes (course, college, family income, etc.) as the independent variables. This regression seeks to predict how the students would rate the quotas conditional on their socioeconomic background. What we find in the data is that 57.7% of the variation found in the student responses can be explained by their socio-economic profile. However, this data needs to be seen as speculative because its 'F-value' is affected by the sample size which was not significant. (see the ANOVA in the appendix - Figure # 5). Still, a similar discourse has been observed by (Bourdieu, 2007) as he summarizes his quota research stating - "these [quota] students do not choose, they are those chosen to occupy careers and courses of lesser prestige."

Multiple Regression Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.760[a]	.577	.023	1.082

a. Predictors: (Constant), School, Course, Race, Gender, Income, Age, Quota Student, Entry Year

b. Dependent Variable: Quota Rating

Table 3: Summary of the Multiple Regression Model. Predictors (constant) are School, Couse, Race, Gender, Income, Age, Quota Student, and Entry Year. The dependent variable is Quota Rating.

Qualitative Results and Discussion

To answer the following short-response question "describe your college experience?," we use data collected on a scale of 1 to 5 whereby students would rank the most present challenges in their experiences. Of the 41 participants, 24 answered the first question and 29 answered the follow-up question. What the responses reveal is simply a division between quota-holders and non-quota holders. For the quota holders, the most pressing challenges they detail (in descending order) were: cost of materials/food/rent/ transportation, balancing working hours with class (many of them take night classes and work during the day), and social exclusion. In other words, most of their problems posed literal consequences for their attendance in the university. On the other hand, the challenges described by non-quota holders were not related to their permanence in the institution (e.g. securing an internship, participating in extracurricular, accessing to teachers, transferring what they learn in the classroom into tangible skills). It is interesting to note that the challenges reported by the students in the state of Amazonas were outliers. For these students, both quota holders and non-quota holders alike described facing a lack of infrastructure on their campus. This may suggest that the students' problems are not centralized, but rather, that the challenges are nuanced and move depending on the region.

In addition, several quota students shared in the interviews a feeling of "happiness to simply attend a public university", as they view it as a life-changing opportunity to improve their socioeconomic status. For many of these students, studying at the public university is the most constructive option in their lives. However, this attitude may come with a conundrum of encouraging these students to apply for a place in the public colleges regardless of their course of study. This may explain the tendency for quota students to look for the least competitive courses in order to better their chances of acceptance into a free institution. If these assumptions are valid, they would contribute to an internalized discourse on race and socioeconomic status that dictates that underrepresented students attend whichever courses are offered at night, whichever courses require the least materials (to minimize expenditures), and whichever course they have the best chance of gaining acceptance. In this case, if quotas are unsuccessful in changing these attitudes, then they are not effectively solving the problem of unequal access to educational resources.

Conclusion

The proposal of this text was not to validate or invalidate the presence of quotas but rather to examine using a microscopic-lens how students interact with the quotas, to potentially justify alternatives or alterations to existing policy. Consequently, this study was designed to access the beliefs and attitudes of Brazilian students about their experience in public as it relates to AA policy. Based on the collected *qualitative* data of the student narratives in interviews and questionnaires, we confirm evidence of an internalized discourse on race and socioeconomic status that dictates how these factors interact with students' access to higher education. This discourse associates the most competitive courses of greater prestige (such as law or business administration) with higher family income. Indeed, our qualitative data implicates the presence of this narrative positively disproportionately correFlating underrepresented

students with less prestigious institutions/course selection. It may be that underrepresented students are subject to a selection bias whereby they are selected by courses that are less competitive and have a lower cost of materials.

Furthermore, from this small, random, sample size we can glean that the general perception of Brazilian AA policy is that it is ineffective. We have seen nuances in student accounts of the challenges they face on their college campuses and these challenges move depending on the region, this perhaps means that the problems of Brazil's education system operate on various geographic and governmental levels. More, what we can deduce from this point is that simply reserving seats for underrepresented students without other measures is not enough to reduce inequalities in the education system. Preparing underrepresented students to receive higher level education is another important part of closing the education gap. One step in particular to do this would involve investing in the lower levels of the public school system (e.g. high schools) that most underrepresented applicants come from.

Solving inequality in the education system is not only a matter of distributing seats, but also, financial management is a factor that directly affects underrepresented students. For this reason, to address the challenges mentioned in the Brazilian education system, I propose a three- point plan that requires greater transparency in the educational budget, followed by a reassessment of the distribution of educational funds by the Ministry of Education and the implementation of a new tuition-system for Public Universities, in which students pay tuition on a sliding percentage scale based on their family income or or attend free of cost if they are below a certain income level. With the remaining savings from the new tuition based school system, I would recommend reinvestment into the country's primary and secondary education system. I would also submit that it is important to start campaigning to change the narrative on college campuses surrounding AA policy so that these policies can be better understood as inclusive efforts rather than alienating. Everyone stands to benefit from attending a more diverse campus[12]. However, when quotas are casted as a socio-economic/racialized competition, underrepresented students face large social costs which may deter them from applying to particular courses. When viewed as a cut-and-dry competition, AA policy appears to polarize campuses over issues of race, class, and who does or does not deserve to attend. For this reason, the status-quo on many Brazilian campuses is tangibly hostile towards quota holding students. In the future, I propose studying the trajectory of these students, investigating their attrition rates in their programs of study to see if the cause is due to lack of scholarships/aid, lack of night school offerings, etc. I would additionally obtain a larger sample to gain better statistical analysis and further confirm or refute the notion that quota students actually face more "bullying" for their status.

References

1. Beting, Joelmir. "Canal Livre Discute Cotas Raciais." Canal Livre, Rede Bandeirantes, São Paulo, São Paulo, 10 Apr. 2009.
2. Soares, Maria Susana Arrosa., and Arabela Campos. Oliven. "Educação Superior No Brasil." Educação Superior No Brasil, CAPES, 2002, pp. 24–37.
3. Soares, Maria Susana Arrosa., and Arabela Campos. Oliven., pp. 24–37.
4. Soares, Maria Susana Arrosa., and Arabela Campos. Oliven. "Educação Superior No Brasil." Educação Superior No Brasil, CAPES, 2002, pp. 194–218.
5. Soares, Maria Susana Arrosa., and Arabela Campos. Oliven. "Educação Superior No Brasil." Educação Superior No Brasil, CAPES, 2002, pp. 113–130.
6. Instituto Nacional de Estudos e Pesquisas (INEP) Educacionais Anísio Teixeira. Censo Da Educação Superior 2013, Ministério Da Educação, 2013, download.inep.gov.br/educacao_superior/censo_superior/apresentacao/2014 coletiva_censo_superior_2013.pdf.
7. INEP. Censo Da Educação Superior 2013.
8. Beting, Joelmir. "Canal Livre Discute Cotas Raciais."
9. Educarargentina, director. Raça Humana. Raça Humana, YouTube, 29 Oct. 2010, www.youtube.com/watch?v=ovZVqvkyBbo.
10. Beting, Joelmir. "Canal Livre Discute Cotas Raciais."
11. Vermelho, Portal. "Indígenas Da UFSC Repudiam Agressão a Cotista Kaingang No Sul." Portal Vermelho, Portal Vermelho, 29 Mar. 2016, www.vermelho.org.br/noticia/278475-10.
12. Gurin, Patricia, Biren Ratnesh A. Nagda, and Gretchen E. Lopez. "The benefits of diversity in education for democratic citizenship." Journal of social issues 60.1 (2004): 17-34.

Appendix

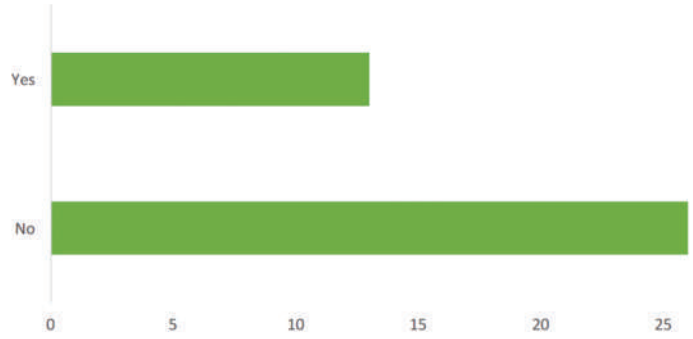


Figure 1: Was the participant admitted as a quota holder (*cotista*)?

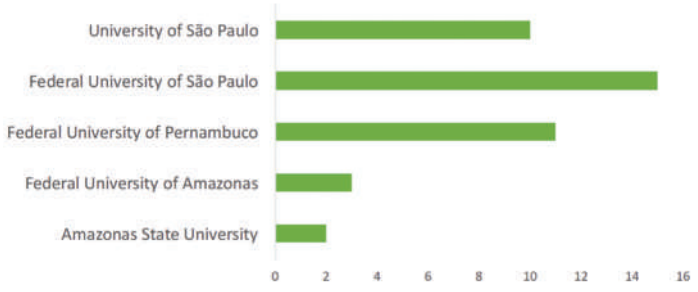


Figure 2: Distribution of Participant Universities

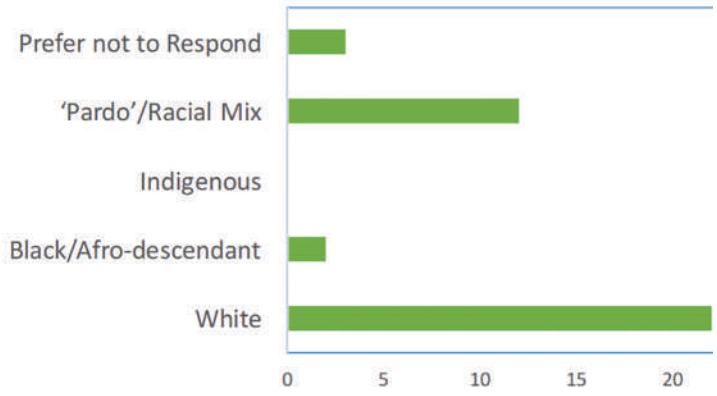


Figure 3: Racial Distribution of the Participants

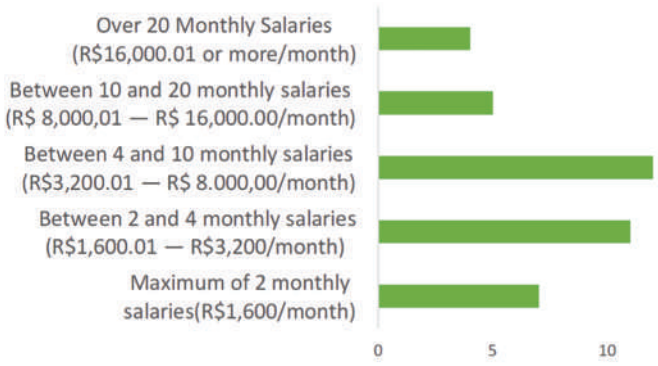


Figure 4: Distribution of Participants' Household Income

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	25.605	21	1.219	1.041	.475[b]
	Residual	18.737	16	1.171		
	Total	44.342	37			

a. Dependent Variable: QuotaRating

b. Predictors: (Constant), School=Universidade do Estado do Amazonas, Race=Branco/europeu, Course=Social Service, Course=Osasco - SP, Course=Accounting, Course=Recife, Course= Social Science, Course=Pedagogia, Course=Direito, Race=Negro/afro-decendente, Course=Geografia, Age, Course=History, School=Universidade de SP, Masculino, Race=Prefer not to answer, Course=Turismo, QuotaStudent, HouseholdIncome, EntryYear, School=Universidade Federal de SP

Figure 5: Multi-variable regression with students' ratings of the quota system as a dependent variable and their attributes (course, college, family income, etc.) as the independent variables.

Pursuing Synergy: Combining a Pan-Class PI3K Inhibitor with Novel Small Molecule Inhibitors of MYC

Muhammed Ors '17

One of the most common aberrations in human cancers is the overexpression of MYC, a master regulator transcription factor which functions in a heterodimer with the MYC-associated protein X (MAX) (Arvanitis & Felsher, 2006). Previously, it was shown that transduction of MYC caused resistance to PI3K inhibitor GDC-0941 (Muellner et al., 2011), and that the drug JQ1 inhibited MYC expression-induced PI3K rescue through JQ1's inhibitory effects on the protein BRD4, which functions as a MYC enhancer (Stratikopoulos et al., 2015). Here we use novel small molecule modulators of MYC transcriptional activity, KI-MS1- 001 and KI-MS2-008, in a comparative study across three human cell lines against the aforementioned PI3K inhibitor GDC-0941 and BET inhibitor JQ1. Co-immunoprecipitation experiments between purified Streptavidin Binding Protein (SBP) tagged MYC, the SBP tag being used to purify the MYC from whole cell lysate, and pure MAX protein in the presence and absence of the compounds imply that the novel MAX modulator KI-MS2-008 does not disrupt the MYC/MAX heterodimer binding, but data also seemed inconclusive and more work needed to be done for a definitive answer. Cell viability assays in the breast lines HCC1599 and MDA-MB-468 confirms that KI-MS1-001 and KI-MS2-008 (in addition to JQ1 and GDC-0941) reduces the viability of these cell lines, more so than these modulators have in previous studies done by the Koehler lab in other lines, but the brain line U87MG showed non convergent IC50 values for all compounds. Experiments are currently being done to determine if the compounds lower MYC and MAX protein levels in a dose dependent fashion in these cell lines. While this work does not rigorously support the hypothesis that the novel modulators would perform better than the other compounds, it contributes preliminary data for any future work the Koehler Lab will do that could more thoroughly test this hypothesis and optimize the modulators.

A transcription factor is a protein that can bind to DNA to help control the rate of transcription. Dysregulated transcription factors can lead to unregulated cell growth and gene expression and therefore is a key therapeutic target in cancer research (Darnell, 2002). Transcription factors as a class of therapeutic targets that would be both unique enough that drugs could be designed specifically to target the transcription factor of interest, and broad enough that treatment techniques would be more than just a single-disease solution.

MYC is a master regulator transcription factor due to its pervasive necessity in transcriptional activity; MYC is part of the MYC/MAX/MAD family of basic helix-loop-helix leucine zipper (bHLHZip) domain transcription factors (Adhikary & Eilers, 2005). It functions in a heterodimer with the protein MAX, the two intrinsically disordered individual proteins undergoing a coupled folding and binding to DNA and a number of additional cofactors in order to then either activate or repress transcription (Adhikary & Eilers, 2005). As demonstrated both in Burkitt's lymphoma and in a study of the transcriptional network of Drosophila genome, the MYC/MAX heterodimer occupies approximately 15% of gene promoters in cells, healthy and cancerous (Li et al., 2003; Orian et al., 2003). These occupied genes are incredibly diverse: MYC regulates cellular processes such as angiogenesis, apoptosis, differentiation, growth, metabolism, proliferation, protein synthesis, ribosomal biogenesis, and self-renewal (Adhikary & Eilers, 2005; Dang, 1999; van Riggelen, Yetil, & Felsher, 2010). In any one of these fields, MYC would be taken seriously as an important protein. As a result, the overexpression of MYC is one of the most common aberrations in human cancers, having been linked to upwards of 70% of cancers (Dang, 2012). Naturally, these aberrations range greatly from uncontrolled cell proliferation and immortalization to escape from immune surveillance and growth factor independence (Vita & Henriksson, 2006). Current estimates directly attribute 100,000 annual cancer deaths in the U.S. to the deregulation of MYC activity (Boxer, 2001).

One important concern when considering to target MYC as a cancer therapy is that for the same reason that inhibiting dysregulated MYC would be widely applicable and desirable, it could also cause dangerous toxicity to the healthy cells that rely on MYC activity for normal processes (Soucek et al., 2008). Fortunately, research has shown inactivation of MYC leads to tumor regression and apoptosis, and systemic effects of MYC inhibition are well tolerated and are completely reversible (Arvanitis & Felsher, 2006; Soucek et al., 2008). Furthermore, in osteogenic sarcoma even brief inactivation of MYC was sufficient, and reactivation of MYC did not restore the cancer (Jain et al., 2002). An important consideration for why systemic MYC inhibition may not be as harmful as once feared is because at any given time the normal cells in the body are quiescent and do not express much MYC anyway; side effects may not be any worse than side effects occurring in hematopoietic and gastrointestinal tissues – cells that are known to replicate often – due to current medicines (Prochownik & Vogt, 2010).

While effects of MYC inhibition appear to be tolerable for normal cells, this is decisively not the case for cancer cells where dysregulation of MYC is a big driver in their tumorigenesis. In these cases, cancers ranging from lymphoma, leukemia, osteosarcoma, and many more, inactivation of the MYC pathway alone can lead to sustained tumor regression through proliferative arrest, apoptosis, senescence, etc. (Felsher, 2010). This phenomenon is referred to as oncogene addiction: when a tumor heavily dependent on a single oncogenic pathway dies due to inhibition of that pathway.

Therefore, it is clear that identifying a small molecule that disrupts the MYC/MAX heterodimer or otherwise modulates MYC function in vivo can have immense implications in drug design, cancer treatment, and survival rates. Unfortunately, MYC historically being considered "undruggable" due to lack of typical binding pockets has stunted research on it (Darnell, 2002). However, this moniker may be more outdated than accurate: research on a variant ETS transcription factor (ETV1) that undergoes undergoes chromosomal translocation in prostate cancers and Ewing sarcomas, among other conditions, using small-molecule microarray screens found a drug-like compound named BRD32048 that binds ETV1 directly and modulates its transcriptional activity (Pop et al., 2014). Simply considering oncogenic transcription factors "undruggable" is no longer good enough.

MYC Modulators Past and Present

With the goal of finding inhibitors of the MYC/MAX heterodimer, Yin, Giap, Lazo and Prochownik screened a chemical library of 10,000 low molecular-weight compounds using a yeast two-hybrid system in which disruption of the MYC/MAX heterodimer led to a decrease in beta-galactosidase activity (Yin, Giap, Lazo, & Prochownik, 2003). From this, seven compounds were found to be specific inhibitors of the MYC/MAX heterodimer in a control yeast strain (Yin et al., 2003). Compounds from this initial set have since been used as positive controls in inhibition experiments, although synthetic chemistry work has been done to try and increase their potency; one compound in particular, 10058-F4 (IC50 = 49 μM, in HL60 cells), can inhibit growth of fibroblasts, but improved synthetic chemistry efforts from the same lab still only led an increases in potency to approximately IC50 = 20 μM (Wang et al., 2007).

The Koehler Lab has been working to discover and develop new probes with better potency or selectivity over these existing MYC/MAX modulators. Previously, Bradner, McPherson, and Koehler developed small-molecule microarrays (SMMs), glass microscope slides onto which solid substrates have been arrayed with the purpose of identifying potential small molecule modulators for transcription factors (Bradner, McPherson, & Koehler, 2006). These SMMs were developed using a novel isocyanate-based covalent attachment technique which allows for the inclusion of compounds like bioactive small molecules, FDA-approved drugs, synthetic drug-like compounds, and natural products into screens—compounds which had not been intentionally synthesized to be arrayed onto a solid substrate, but because they had certain functional groups on them, like thiols, primary alcohols, and amines, the researchers were able to adhere them to the slides (Bradner et al., 2006). Following the attachment of compounds, SMMs are incubated with pure protein of interest or cellular lysate, with protein-small molecule interactions being detected through fluorescence (Bradner et al., 2006). Using this novel SMM development method, Clemons et al. performed a screen of more than 20,000 compounds against 100 transcription factor proteins, including MYC and MAX, and identified potential small molecule binders to MYC and MAX (Clemons et al., 2010). Koehler then performed another screen against pure MYC and pure MAX of more than 40,000 compounds, including ones previously identified in the Clemons et al. screen. It is from here that the prospective MYC/MAX modulators KI-MS1 and KI-MS2 emerged, and this is currently being written into another paper.

The goal of this unbiased-binding approach was to find novel direct probes of MYC independent of it being bound to MAX, for this could provide new information to the field as all known direct probes of MYC have been found through binding against the MYC/MAX heterodimer. The 313 positive hits discovered through the SMMs were subsequently evaluated in secondary reporter gene assays in HEK293T cells; 32 of these compounds specifically inhibited MYC-dependent transcription. Currently, the Koehler Lab is characterizing the mechanisms of action of the most promising of these hit compounds using cell-free and cell- based approaches, and further chemically optimize their potency and selectivity. An ultimate goal is to use these putative MYC modulators in translational studies involving actual cancers where dysregulated MYC has a large impact and thus break the ground around exploring therapeutic modulation of an "undruggable" oncogenic transcription factors like MYC.

MYC's Role in the PI3K/AKT pathway

While inhibition of proteins and enzymes with roles in important kinase pathways provide attractive approaches to cancer treatment, sustaining this inhibition has proven to be a major challenge. One such pathway that develops inhibitor resistance is the PI3K/AKT/mTOR pathway, which is centered on phosphoinositide-3 kinase (PI3K) and is important in regulating processes involved in the cell cycle, proliferation, and longevity (Figure 1). Relevantly, MYC has been proven to be key player in it: during a study on the PI3K pathway using PIK3CAH1047R- initiated mammary tumors Liu et al. found that the tumors able to escape oncogene addiction had elevated levels of MYC, with knockdown of MYC reducing the incidence of recurrent tumors (Liu et al., 2011). Muellner et al. similarly observed that MYC mRNA translation and MYC transcriptional activity were amplified more often among human cell lines resistant to PI3K- mTOR inhibitors, and subsequently demonstrated that resistance to PI3K inhibitors was dependent on MYC induction (Muellner et al., 2011). Furthermore, RNA interference (RNAi) of MYC mRNA reversed the observed inhibitor resistance, showing its necessity and sufficiency for PI3K-mTOR resistance (Muellner et al., 2011).

Inspired by these results, Stratikopoulus et al. pursued a strategy to overcome PI3K inhibitor resistance using a metastatic breast cancer mouse model driven primarily by proteins PI3K and MYC. To make this model they crossed mice strains that expressed MYC mutations to strains that had either PIK3CA overexpression (H1047R mutation) or loss of PTEN expression (Ptenflox/flox) to create strains that overexpressed both MYC and PI3K, referred to as PI3K;MYC mice (Stratikopoulos et al., 2015). This model resisted a pan-class 1 PI3K inhibitor that hit both p110α and p110β subunits (GDC-0941) through feedback activation of tyrosine kinase receptors (RTKs), AKT, mTOR, and MYC. Another experiment showed their mouse model was also resistant to inhibitors of bromodomain and extra terminal domain (BET) proteins, which recognize acetylated-lysine residues in nucleosomal histones and facilitate recruitment of transcription proteins to chromatin. Stratikopoulos et al. chose to use the BET inhibitor JQ1 in order to target a member of the BET protein family BRD4, due to the fact that JQ1 had been previously shown to suppress MYC activity by downregulating MYC transcription and thereby causing potent anti-proliferative effects associated with cell cycle arrest and cellular senescence (Delmore et al., 2011; Stratikopoulos et al., 2015). Stratikopoulus et al. found

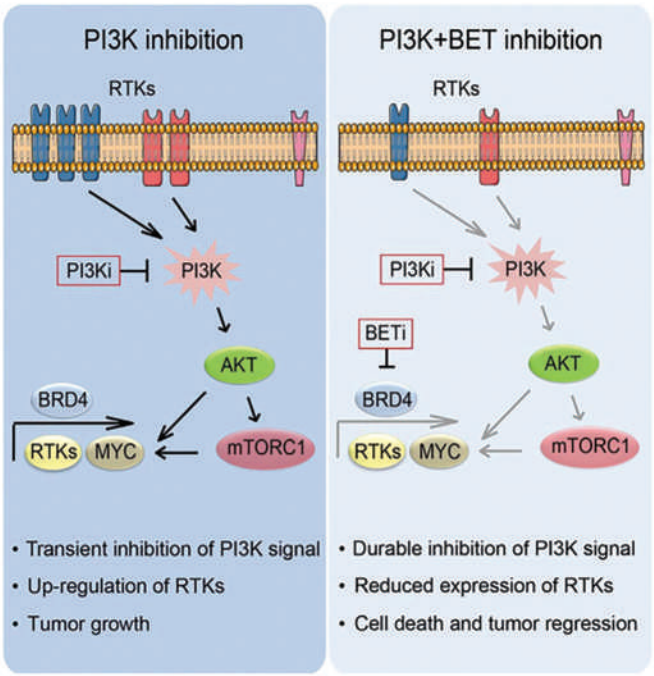


Figure 1. Graphical Representation of the PI3K/AKT/mTOR pathway. The left half illustrates that when there is only PI3K inhibition, PI3K is rescued from this inhibition due to the downstream function of AKT and mTORC1 in activating MYC and causing a feedback loop. The right half illustrates the prevention of this rescue when there is simultaneous PI3K and BET inhibition, due to the prevention of MYC transcription. (adapted from Stratikopoulos et al., 2015).

that BET and PI3K inhibitors combined led to a sustained decrease of signaling of PI3K and proteins downstream of it in cell lines from the PI3K;MYC mice, causing cancer death and tumor regression both in vitro and in vivo (Stratikopoulos et al., 2015). They believe these results provide an appropriate step forward in pursuit of preventing the rescue of PI3K inhibition (Stratikopoulos et al., 2015).

Project and Hypothesis

This study seeks to capitalize on the preliminary data and approach communicated by Muellner et al. and Stratikopoulos et al. by performing an in vitro synergistic study of PI3K and BET inhibitors—but with the additional dimension of directly modulating MYC activity using the novel compounds of the Koehler Lab. The cell lines used include commercially available human cell lines corresponding to two breast cancers, HCC1599 and MDA-MB-468, and one brain cancer, U87MG. The latter two were chosen due to the loss of PTEN expression and/or mutation in PIK3CA which leads to overexpression of the PI3K pathway, and because both were present in the Stratikopoulos et al. study. HCC1559 will be a control line which does not have either loss of PTEN or a PIK3CA mutation. The compounds used are JQ1 and GDC-0941 from the previous study, and the Koehler Lab's KI-MS1-001 and KI-MS2-008. The first aim is to determine if the MYC modulators disrupt the MYC/MAX heterodimer. The second aim is to characterize the independent IC50s of the four inhibitor compounds using cell viability assays, followed by western blotting to look for any dose-dependent decrease in MYC and p-AKT protein following incubation. The third aim, which was not completed during the thesis project period, involves determining the IC50s of the synergistic combinations of the PI3K/BET and PI3K/c- MYC inhibitors. The hypothesis is that the synergistic effects of the PI3K/MYC inhibitors will produce a lower IC50 than those of the PI3K/BET inhibitors.

Results

Purification of SBP Tagged MYC

We grew bacteria expressing high amounts of SBP-MYC by using a frozen glycerol stock of E. coli bacteria which had previously been transformed by Dr. Caballero with a MYC vector pET28-6xHis-SBP-TEV (Addgene). After separating the SBP-MYC soluble lysate from the insoluble pellet, my next goal was to purify the SBP-MYC from the rest of the proteins in the cell lysate, and also the MYC protein from the SBP tag. I performed a Coomassie stain and a western blot to check the efficacy of the purification approach. The purified MYC would be found in the sample named FT-TEV on the blot, the supernatant separated from the streptavidin agarose beads after incubation with the TEV Reaction Mix (Figure 2a). A sample of the flow through after the overnight incubation was saved and named FT-Beads as a way to provide a comparison point to the purity of the isolated SBP-MYC (Figure 2a). The FT-Beads sample would be used to demonstrate just how much of the SBP-MYC is left unbound to the Streptavidin agarose beads after incubation, though admittedly this is perhaps not a positive or negative control exactly—I reasoned that it would provide an internal control of the

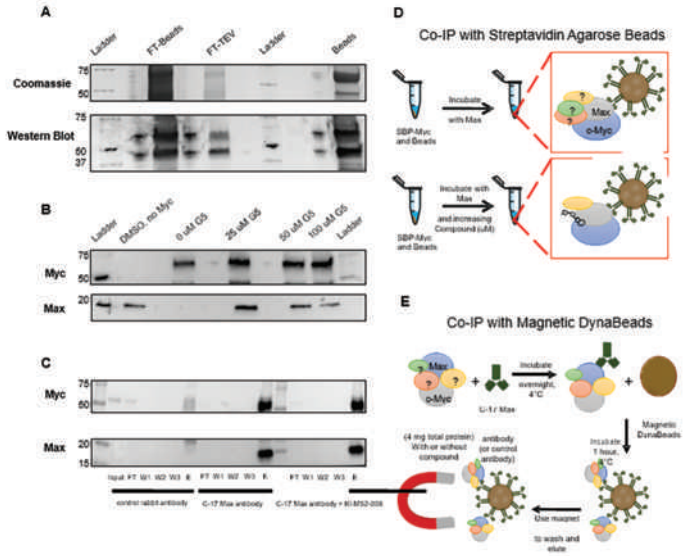


Figure 2: KI-MS2-008 does not disrupt the MYC/Max heterodimer. (A) SBP- MYC was purified from E. coli. bacteria using Streptavidin agarose beads, and then the SBP tag was further purified from the MYC using AcTEV Protease, but the yield seemed low using these conditions. (B) Co-Immunoprecipitation of Max using SBP-MYC in presence of increasing concentrations of positive control 10074-G5 seems to show that 10074-G5 inhibits the heterodimer binding, but the control did not work so it is hard to conclude decisively. (C) Co-Immunoprecipitation of MYC and Max using anti-Max (C-17) antibody, with and without novel Max modulator KI-MS2-008, though KI-MS2-008 does not seem to inhibit the heterodimer. Lanes are labelled Flowthrough (FT), Wash (W), Elution (E) (performed by Andrew Chen). (D) Cartoon schematic illustrating the protocol that produced Figure 2b (adapted from Andrew Chen). (E) Cartoon schematic illustrating the protocol that produced Figure 2c (adapted from Andrew Chen).

efficacy of incubation conditions. Furthermore, a second internal control was devised to look at the efficacy of the protease reaction by simply adding Laemmli running buffer to the Streptavidin agarose beads following the TEV reaction, creating a sample that would prospectively show what was left attached to the beads and was unable to be isolated through the purification (Figure 2a).

The western blot showed MYC bands around 57 kDa for all three lanes, FT-Beads, FT- TEV, and Beads, which corresponds to the molecular weight of MYC (57-60 kDa) according to anti-MYC antibody manufacturer's information, though the Koehler Lab has many times seen MYC run closer to 50 kDa. The lightest band is seen from the FT-TEV lane, while the darkest band was from the Beads lane (Figure 2a). From these results, the greatest concentration of MYC remained attached to the streptavidin agarose beads instead of being properly cleaved by the AcTEV Protease reaction. In order to produce cleaner and better quality gels to look at I performed these experiments twice, and the second run provided the images used for Figure 2a. However, the results did not change between these two runs and there was not enough time to try again. Considering that cleaving the SBP tag was not absolutely necessary to use the protein, I decided after talking to Dr. Caballero to change the original plan for the rest of the experiments and simply perform them with the SBP-MYC.

Novel Compounds Do Not Disrupt MYC/Max heterodimer

The purpose of these experiments was to ask whether the modulators disrupt the MYC/MAX heterodimer binding. To that end, I chose to look at this question in vitro through co- immunoprecipitation assays between MYC and MAX in the presence and absence of modulator compounds, using the SBP-MYC lysate produced previously as my source of MYC. This first set of assays use the positive MYC/MAX heterodimer inhibitors 10074-G5 or KJ-Pyr-9 as the compounds modulating the interaction between SBP-MYC and MAX; the purpose of this set of assays was to behave as a large control for all of the co-immunoprecipitation assays that were to follow. Within this assay protocol, the negative control was a sample prepared without SBP- MYC, to determine if MAX would be pulled down by the streptavidin agarose beads in the absence of SBP-MYC. The positive control tubes were prepared with increasing concentrations of the MYC/MAX heterodimer inhibitors 10074-G5 (25 μM, 50 μM, and 100 μM) or KJ-Pyr-9 (10 μM, 15 μM, 20 μM). Additionally, all samples were kept at a final concentration of 1% DMSO to be consistent with conditions, as this is what the drug compound stock solutions had been diluted in.

Blots were visualized through chemiluminescence using the ChemiDoc MP Imaging System machine (Bio-Rad) and they were further analyzed using the Image Lab program (Bio- Rad) in order to ascertain the relative intensities of the visible bands (Figure 3). Bands for MYC were seen in all lanes but the lane where there was no SBP-MYC lysate added, as expected (Figure 2b). However, amongst the bands for MAX while there seemed to be a decrease in MAX protein present as 10074-G5 increases in concentration, I was surprised to see a band for MAX coming from the MYC-absent sample and no band from the sample incubated without the inhibitor compound (Figure 2b). The Image Lab program analysis of band intensity confirmed that while the MAX protein concentration was decreasing as 10074-G5 concentration was increasing from 25 μM to 100 μM, as evidenced by the decreasing ratio of MAX/MYC

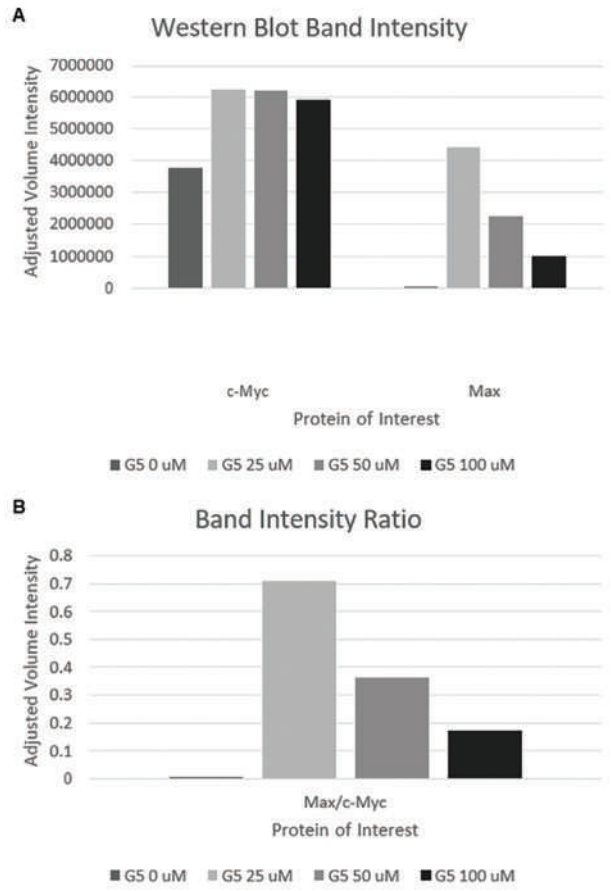


Figure 3: Analysis of band intensity using Image Lab. (A) Higher intensity values were measured for MYC than for MAX. This is most likely due to the fact that the Streptavidin agarose beads are binding to the SBP tag on the MYC, while MAX is only being pulled down through interaction with MYC. There is a decrease in intensity as G5 concentration increases, except for the 0 uM condition. (B) For better visualization of the data, intensity of MYC bands were also compared to intensity of MAX bands in the form of a MAX/MYC ratio. Because 10074-G5 is a MYC/Max heterodimer inhibitor, this ratio would be expected to decrease as the concentration of inhibitor increased due to there being progressively less MAX binding to the MYC. This held true for 25 uM to 100 uM, but not for 0 uM.

present in each lane, this was not the case when the concentration of 10074-G5 was 0 μM (Figure 3b). Considering the unexpected developments encountered here, I performed technical replicates with the same lysate and biological replicates with another round of purified SBP-MYC lysate. However, I was unable to get data that would either replicate what was seen in Figure 2b or contradict it—all of the other gels were very blotchy with undecipherable bands. Therefore, I have chosen to include the clearest looking image in this thesis (Figure 2b).

Following co-immunoprecipitations done with incubation with the positive control compounds, the next step would be to use the small molecule modulators developed in lab, currently designated as KI-MS1-001 and KI-MS2-008. This would have been done to test the hypothesis that these too would inhibit the MYC/MAX heterodimer. Unfortunately, due to time constraints and concerns about moving onto other experiments for the thesis, I could not personally test this hypothesis. Luckily, my mentor Andrew Chen was already working on this question independently and allowed

me to include his results here to serve as a comparison to the previous step (Figure 2c). Within this assay protocol there were two negative controls in a sense: a sample prepared with only DMSO, without either anti-MAX (C-17) antibody or KI-MS2-008, and secondly, for each condition tested three washes were also run as an internal control, as there should have been no proteins of interest visible from these wash steps. The incubation condition of anti-MAX (C-17) antibody without KI-MS2-008 would be the positive control, as the antibody should bind to MAX which would in turn pull down its heterodimeric partner MYC.

The blot was visualized through chemiluminescence using the ChemiDoc MP Imaging System machine (Bio-Rad). No bands were visible for the flow through (1% loaded) or wash (5% loaded) lanes, but there was a faint band for MYC immunoprecipitated from the control sample incubated with normal IgG rabbit antibody—likely due to non-specific binding (Figure 2c). Most importantly, KI-MS2-008 did not seem to significantly impact amount of MAX pulled down by MYC, suggesting it does not interfere with in vitro MYC-MAX binding (Figure 2c).

KI-MS1-001 and KI-MS2-008 Reduce Cancer Cell Viability

The human cancer cell lines used to perform the cell viability assays were two breast cancer cell lines, HCC1599 (suspension) and MDA-MB-468 (adherent), and one brain cell line, U87MG (adherent). U87MG and MDA-MB-468 both have loss of PTEN expression and/or mutations in PIK3CA which leads to overexpression of the PI3K pathway, and additionally both were present in the Stratikopoulos et al. study in a limited manner. HCC1599 has neither loss of PTEN nor a PIK3CA mutation and is meant to be a control line. These cell lines were chosen because they were all readily available from the MIT Koch Institute's high throughput screening facility. The purpose in doing cell viability assays was to characterize the independent IC50s of the pan-class I PI3K inhibitor GDC-0941, the BET inhibitor JQ1, and the two MYC and MAX modulators KI-MS1-001 and KI-MS2-008.

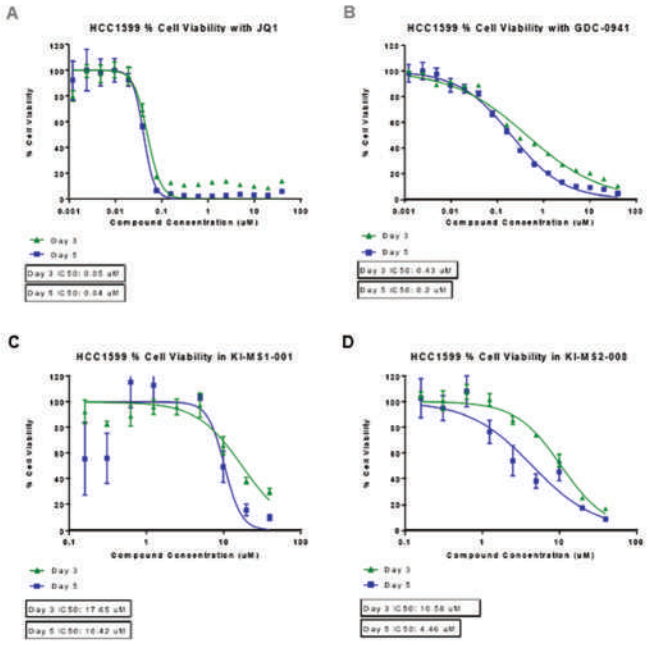


Figure 4: HCC1599 cell viability assay data, second biological replicate. As expected all drug compounds displayed lower IC50 values on Day 5 in comparison to Day 3. (A) All of the drug compounds seemed to be capable of reducing cell viability, with JQ1 to being the most potent out of all four, going so low as 0.04 uM. (B) GDC-0941 also had IC50s less than 1 uM. (C) KI-MS1- 001 had the highest IC50, around 11 uM. (D) KI-MS2-008 had a IC50 around 5 uM.

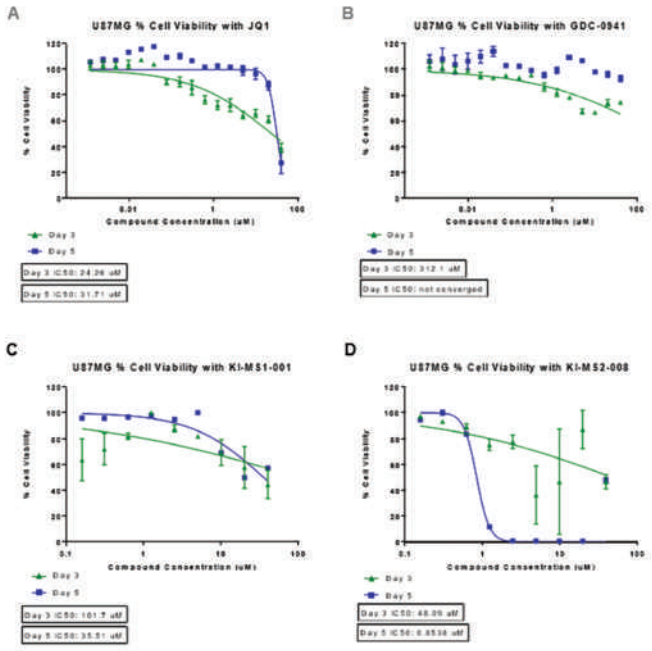


Figure 5: U87MG cell viability assay data, second biological replicate. Overall, it seemed like none of the drug compounds were effective and were delivering IC50 values greater than 30 uM. (A) Unlike any other replicate with any drug in any cell line, here we see that JQ1 had an IC50 value that actually went up from 24 uM on Day 3 to 32 uM Day 5. (B) In this replicate, GDC-0941 was especially ineffective even compared to the other drugs, because the data was unable to converge to an IC50 value for Day 5. (C) The IC50 values for KI-MS1- 001 went down from Day 3 to Day 5 as would be expected, but were still only around 35 uM. (D) Interestingly, KI-MS2-008 had an extremely low Day 5 IC50, around 0.85 uM, which is probably due to error considering the other drugs.

In the first biological replicate all four drug compounds were serially diluted two-fold in DMSO from 40 μM to 0.156 μM (results not included in this manuscript). The second and third replicates were further serially diluted from 40 μM to 0.00122 μM due to a desire to better visualize a sinusoidal curve to calculate IC50s from. Cell lines incubated in compounds or DMSO and for both 3 days and 5 days, after which that plate would be visualized. These time points parallel those tested in the Stratikopoulos et al. paper. CellTiter-Glo was used to measure the amount of ATP present in the wells of the plates through luminescence, as these values are considered to be a standard indicator of metabolically active cells, and therefore a reasonable method to draw conclusions on cell viability (Riss, Moravec, & Niles, 2011). Luminescence values were graphed, and IC50 values calculated, using Prism software (Graphpad). I performed three biological replicates and three technical replicates for each of the conditions tested. The technical replicates are accounted for through the error bars appearing on the graphs, and I have chosen to include both the second and third replicates as they each give interesting and partially conflicting information

about the cell lines.

In the HCC1599 cell line all of the drug compounds seemed to be capable of reducing cell viability, with JQ1 (Figures 4a and S4a) and GDC-0941 (Figures 4b and S4b) consistently having IC50s less than 1 μM, while KI-MS1-001 (Figures 4c and S4c) and KI-MS2-008 (Figures 4d and S4d) had higher IC50s, between 4 μM and 14 μM on Day 5. All drug compounds displayed lower IC50 values on Day 5 in comparison to Day 3, which would be expected; JQ1 seemed to be the most potent out of all four, going as low as 0.002 μM (Figure S4a). In the U87MG cell line, none of the drug compounds were particularly effective, with almost all of the IC50 values for the drugs being greater than 30 μM (Figures 5 and S5), and JQ1 even having an IC50 value that actually went up from 24 μM on Day 3 to 32 μM Day 5 (Figure 5a). Interestingly, in the third biological replicate while the Day 5 IC50 values for JQ1, GDC-0941, and KI-MS2- 008 (Figures S5a, S5b, S5d) could not be determined because the graphs did not converge, KI- MS1-001 alone had an IC50 value of 3.7 μM at Day 5 (Figure S5c). The data from the MDA- MB-468 cell line were especially interesting as KI-MS1-001 and KI-MS2-008 showed very different IC50s between the second and third biological replicates. All drug compounds had IC50s around 1 μM in the second replicate (Figure 6) but in the third replicate KI-MS1-001 went up to 5.5 μM (Figure S6c) and KI-MS2-008 jumped up to 30.5 μM (Figure S6d).

Following testing of cell viability under compound and DMSO conditions, HCC1599 and MDA-MB-468 were also prepared for western blotting against MYC, p-AKT, and GAPDH in order to test for potential dose-dependent differences in protein expression. I chose to incubate the cells for 24 hours because that was the time frame the Stratikopoulus et al. paper used and I wanted to be as comparable as possible. I chose to incubate with two-fold serial dilutions from 10 μM to 0.156 μM because the measured IC50s seemed to consistently fall within that range. Total protein concentration was determined using a BCA assay in order to be able to prepare samples for the gel to have a consistent 30 μg concentration of protein, otherwise the comparison of bands would not be particularly meaningful. However, by the time the project ended I had been unable to produce any western blot data that showing the MYC protein present in the lysate.

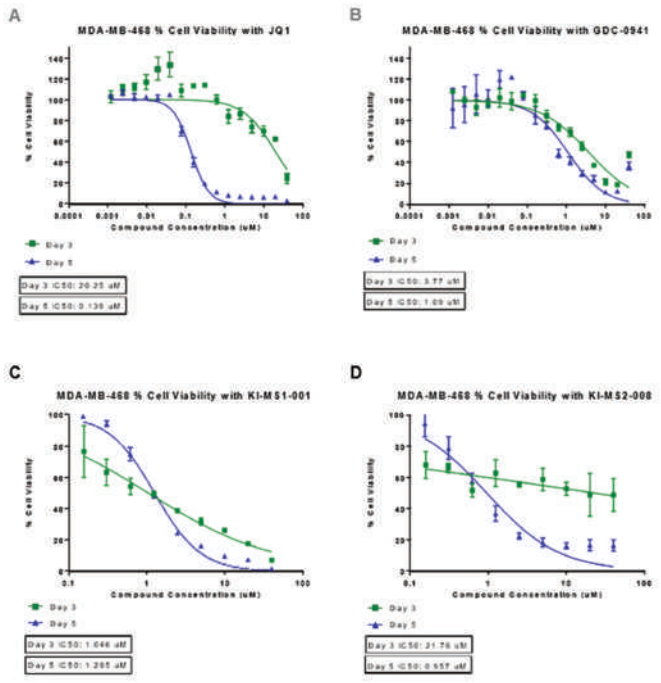


Figure 6: MDA-MB-468 cell viability assay data, second biological replicate. (A, B, C, D) Like the second replicate, all drug compounds again displayed lower IC50 values on Day 5 in comparison to Day 3. All drug compounds had IC50s around 1 uM, an interesting result considering the other cell lines did not show such consistency amongst all four drugs.

Discussion

MYC is indisputably a genetic heavy weight. Studies in both Burkitt's lymphoma and the Drosophila genome have shown that the MYC/MAX heterodimer occupies approximately 15% of gene promoters in cells (Li et al., 2003; Orian et al., 2003), affecting processes including angiogenesis, apoptosis, differentiation, protein synthesis, and many others (Adhikary & Eilers, 2005; Dang, 1999; van Riggelen et al., 2010). Furthermore, overexpression of MYC has been linked to upwards of 70% of cancers (Dang, 2012) with estimated 100,000 annual U.S. cancer deaths directly caused due to deregulation of MYC activity (Boxer, 2001). Most importantly, we know that inactivation of MYC can lead to tumor regression and apoptosis without adversely affecting normal tissue (Arvanitis & Felsher, 2006; Soucek et al., 2008). It seems abundantly clear that identifying a small molecule disrupting the MYC/MAX heterodimer or modulating c- MYC function in vivo should be very high on the list of anyone who wants to "cure" cancer.

The fact that MYC has been considered "undruggable" for so long due to its lack of typical binding pockets is a golden opportunity—whoever creates a viable treatment targeting the protein will surely achieve greatness. It is through the Koehler Lab, and the novel small molecule MYC and MAX modulators being developed in-house, that I have had the opportunity to contribute to this cause. When considering my own thesis I was inspired by the papers written by Muellner et al. and Stratikopoulos et al. examining inhibitor resistance in the PI3K/AKT/mTOR pathway. Muellner et al. had demonstrated that resistance to PI3K inhibitors was dependent on c- MYC induction, which was reversible through using RNA interference (RNAi) of MYC mRNA (Muellner et al., 2011). Stratikopoulus et al. examined PI3K inhibitor resistance using a metastatic breast cancer mouse model driven primarily by either PIK3CA overexpression or loss of PTEN, finding it was resistant to both a pan-class 1 PI3K inhibitor GDC-0941 and BET inhibitor JQ1, but not both together (Stratikopoulos et al., 2015). Looking at these results I had to ask: what about a combinatorial study using the PI3K inhibitor GDC-0941 in conjunction with the Koehler Lab's MYC modulators instead of the BET inhibitor JQ1? KI-MS1-001 and KI- MS2-008 would affect MYC/MAX's transcriptional ability while JQ1 would affect the ability for MYC to even be transcribed, so I thought there might perhaps be interesting things to be learned from comparing these two approaches.

As the first step in pursuit of my thesis, I wanted to examine whether the novel small molecule modulators behaved as heterodimer inhibitors. This was an ongoing question in the Koehler Lab at the time, so that we better understood the mechanism of action of these modulators to put their effects in the cell lines in context. I began this step by growing E. coli bacteria overexpressing SBP-tagged MYC protein and working to purify the protein from the rest of the lysate, performing a western blot to confirm MYC was being

produced and was purified properly. However, as mentioned earlier, the lightest band is seen from the FT-TEV lane while the darkest band was from the Beads lane, the opposite of what I would have wanted (Figure 2a). Although I was able to purify MYC from the rest of the lysate and the SBP-tag as determined from the presence of the FT-TEV band, an ideal purification would have had a much stronger band for MYC in that lane, and weaker bands in both of the other lanes. Especially the Beads lane made from boiling the leftover Streptavidin agarose beads in Laemmli running buffer, as the only thing remaining on these beads should have been the SBP-TEV tag and not MYC.

One explanation why the western blots were against expectations could be because the AcTEV Protease was older than believed and degrading, and thus wouldn't be able to catalyze the cleavage reaction as well explaining the little MYC in the FT-TEV. Another reason could be because I changed the protocol after the first attempt to purify came out very poorly from the optimal incubation conditions of 3 hours in a 37° C water bath to an overnight incubation at 4 ° C. Furthermore, I didn't think to wash the beads after removing the FT-TEV and perhaps there was a lot of cleaved MYC that was just stuck in the beads considering MYC is a very sticky protein. Repeating these purification experiments with new AcTEV Protease would be appropriate, but I decided at Dr. Caballero's recommendation to simply continue to use the SBP- MYC in further experiments as ultimately the SBP tag is small and would not interfere with the MYC/MAX binding interaction. In addition, it is clear that there was overflow in the lanes of the western blot due to my error in loading them, and a repeated effort should load less sample. Any future attempt to purify SBP-MYC would take these issues into consideration.

To examine whether KI-MS1-001 or KI-MS2-008 inhibited the MYC/MAX heterodimer I first performed co-immunoprecipitation experiments between purified SBP-MYC and pure MAX protein in the presence and absence of the drug compound 10074-G5 to establish data for a positive control (Figure 2b). I expected that increasing concentration of the positive control would lead to decreasing concentration of MAX binding to the SBP-MYC, but this turned out to be harder to prove than anyone had reasonably considered. Generally, an increasing concentration of 10074-G5 caused a progressively lesser concentration of MAX to bind to the SBP-MYC (Figure 3). Unfortunately, there was an extremely small amount of MAX binding to the SBP-MYC in the control sample to which no 10074-G5 was added (Figure 3). Considering that the control within this blot clearly didn't work, I performed three more co-Immunoprecipitations with 10074-G5 and two with PY9, another positive control inhibitor—and yet, all of these blots were very blotchy so I was still unable to either replicate what was seen in Figure 2b or contradict it.

There are a couple of reasons that could explain why these western blots were unsuccessful. One reason could be that 100 µL of streptavidin agarose beads was insufficient to incubate with, although this amount had been standard in the lab. Another problem could have been that incubating the five samples for 1.5 hours at room temperature, prior to adding MAX and after, was not long enough for meaningful binding. A third reason for these difficulties could be due to the binding buffer solution used in incubation steps being a buffer solution from an EMSA kit, potentially causing issues in a co-immunoprecipitation; the choice to use this buffer solution had been made because the original plan was to later perform an EMSA testing whether the positive control inhibitor, KI-MS1-001, or KI-MS2-008 affected MYC/MAX binding to DNA, and I had been advised to prepare these Western blots using similar conditions to what would come later. This preparation proved completely unnecessary as I did not end up having time to perform an EMSA. Either way, I feel that these positive control data are inconclusive as I was ultimately unable to properly replicate results in readable gels.

Following the co-immunoprecipitation with the positive control inhibitors, the plan was to test whether the novel small molecule modulators inhibit the MYC/MAX heterodimer by incubating a solution of purified MAX and SBP-MYC with increasing concentrations of both KI- MS1-001 and KI-MS2-008. Unfortunately, I was also unable to perform this due to time constraints. Therefore, I have instead included in this study co-immunoprecipitation experiments that my second mentor Andrew Chen performed with KI-MS2-008 to use as a comparison to my experiment (Figure 2c). As expected, no bands were seen at all for the flow through or wash lanes, and there was only a light band in the control lane (Figure 2c). I was a bit surprised to see bands for MYC and MAX coming from the control sample incubated with normal IgG rabbit antibody, but this is most likely due to non-specific binding as MYC is known to be a sticky protein (Figure 2c). Most importantly, KI-MS2-008 did not seem to significantly impact the amount of MAX pulled down by MYC which suggests that it in fact does not interfere with in vitro MYC/MAX binding (Figure 2c). In the years following these experiments the Koehler Lab has also compiled more evidence to support this case.

I recognize that a couple key conditions are different between the two protocol designs and should be addressed, but I do not think any of these distinctions would disqualify the data from being a useful comparison: firstly, my experiment was performed using SBP-MYC purified from E. coli lysate while Andrew Chen's was performed using normal MYC expressed in the Burkitt's lymphoma cell line P493-6; secondly, my experiment used streptavidin agarose beads while Andrew Chen's used magnetic Dynabeads; and thirdly, I incubated using EMSA kit binding buffer and Andrew Chen incubated using the Dynabead kit binding buffer. Given time, I would have liked to perform the ideal co-immunoprecipitation and EMSA experiments with both KI-MS1-001 and KI-MS2-008 under similar conditions to the positive control experiment.

Following this first goal, the second goal of my thesis was to characterize the independent IC50s of the four drug compounds in cell viability assays. Given limited time, I chose cell lines different along multiple axes to amplify information gained i.e. the cell lines that overexpressed the PI3K pathway were in different cancer types, breast (MDA-MB-468) and brain (U87MG), and the third cell line was a control with no overexpression of the PI3K pathway (HCC1599). I reasoned that without hard independent data in these three cell lines with these four drugs (pan-class I PI3K inhibitor GDC-0941, BET inhibitor JQ1, and the two MYC and MAX modulators KI-MS1-001 and KI-MS2-008), it would be unclear later on whether data produced through a characterization of a synergistic drug combination was actually causing synergistic effects or merely additive ones. Due to a desire to gain more data points from which to better visualize a sinusoidal curve and calculate IC50s of the drug compounds, I chose to perform two-fold serial dilutions from 40 µM to .156 µM and then further to 0.00122 µM.

Interestingly, all of the drug compounds seemed to be capable of reducing cell viability in the HCC15599 cell line: JQ1 (Figures

4a and S4a) and GDC-0941 (Figures 4b and S4b) consistently had IC50s less than 1 μM, while KI-MS1-001 (Figures 4c and S4c) and KI-MS2-008 (Figures 4d and S4d) had higher IC50s, between 4 μM and 14 μM. This was against expectations, as the HCC1599 cell line was expressly chosen because it has neither loss of PTEN nor a PIK3CA mutation. Instead, the cell line is known to be negative for expression of Her2- Neu, p53, estrogen receptor (ER), progesterone receptor (PR). Perhaps the results here, that drugs prospectively meant for the PI3K pathway and MYC/MAX modulators affect the viability of triple negative breast cancer, are the most unexpected gains from my thesis and could be further pursued by the Koehler Lab outside of this study. In the U87MG cell line, it did not seem like any of the drug compounds were particularly effective at all, which combined with the HCC1599 data possibly suggests that there is more of an importance placed on the tissue than the specific mutations than I expected (Figures 5 and S5). Finally, the data from the MDA-MB-468 cell line showed internal contradictions, with all drug compounds having IC50s around 1 μM in the second replicate (Figures 6), but KI-MS1-001 and KI-MS2-008 jumping up in the third replicate (Figures S6c and S6d). One commonly used MYC/MAX heterodimer inhibitor compound that was discovered in the same study as the 10074-G5 used in this thesis is 10058-F4 (Yin et al., 2003). This compound initially was recorded by Yin et al. as having an IC50 of 49 μM in HL60 cells, but improved synthetic chemistry efforts from the same lab still only led an increases in potency to approximately IC50 = 20 μM (Wang et al., 2007; Yin et al., 2003). I would want to repeat the cell viability assays a few more times in MDA-MB-468 to look for more consistency in IC50 values due to the internal contradictions present between the replicates, but considering the data from the Wang et al. paper I think KI-MS1-001 and KI-MS2-008 show promise.

During this stage I took care to do both three biological and technical replicates in order to be most thorough, and naturally there were plenty of opportunities for error when collecting data. The potential for human error whenever I was plating drugs, or media, or cells into the hundreds and hundreds of wells over the course of the months is a particularly large problem. A lot of the variation between technical replicates could probably be explained to a deficiency in technique, especially when I was first performing the assays. Performing these experiments with a robotic system would be a good way to avoid this source of error. Additionally, the western blotting to see if there were dose-dependent decreasing concentrations of the MYC and p-AKT proteins as a result of increasing concentrations of the drug compounds also did not work well. This is perhaps due to the HCC1599 and MDA-MB-468 cell lines not producing enough MYC considering that they are not known to overexpress that protein to begin with. In the future it would be prudent to first determine exactly how concentrated the total protein in the lysate from these cell lines must be to get a detectable band in western blotting.

The third aim had been to bring the thesis study to a close by characterizing the IC50s of the synergistic combinations of the PI3K/BET inhibitors and PI3K/MYC inhibitors, finally testing the hypothesis that the synergistic effects of the PI3K and MYC inhibitors would have a stronger impact on the cell lines and produce a lower IC50 value than the effects of the PI3K and BET inhibitors. As should be clear at this point, I unfortunately wasn't actually able to perform either a combinatorial or synergistic study of the PI3K inhibitor and MYC/MAX modulators. Therefore, the future directions are clear. Stratikopoulos et al. had demonstrated the BET inhibitors blocked PI3K inhibition resistance caused by exposure to GDC-0941, and PI3K/BET inhibition blocked AKT reactivation (Stratikopoulos et al., 2015). They showed this through in vitro proliferation assays performed while holding one of the compounds at a constant 1 μM and adding to it a serial dilution from 8 μM to .125 μM of the other compound. The Koehler Lab plans on finishing my line of questioning by performing cell viability assays using these cancer cell lines; however instead of holding one compound at a constant 1 μM the plan is to vary the concentrations of both compounds along a two-fold serial dilution from 40 μM to .156 μM, in order to more precisely define where the best synergy is.

Through these future studies, we aim to present the novel MYC and MAX modulator compounds KI-MS1-001 and KI-MS2-008 as viable options to put into the cancer researcher's toolkit. As dysregulated transcription factors become increasingly attractive targets for cancer research, developing a small molecule modulator of MYC or MAX function in vivo will only increase in importance as well (Darnell, 2002). In particular, evidence demonstrating elevated levels of MYC increases tumor recurrence in PIK3CAH1047R-driven mammary cancers show that MYC is an important oncoprotein that cannot be overlooked even when attempting to study a seemingly unrelated cancer driver (Liu et al., 2011). The clinical significance of MYC cannot be underestimated, especially in consideration of the work done by Stratikopoulus et al. and Liu et al. in showing that a single PI3K inhibitor isn't enough to use as a treatment for a cancer of the PI3K/AKT/mTOR pathway. Even if the planned synergistic study here does not show that the modulators in their current chemical forms produce better results than the combination of JQ1 and GDC-0941, this wouldn't be a reason to give up on this line of reasoning. These compounds are continually being improved and optimized by the Koehler Lab with the belief that the failures of today will inform the successes of tomorrow. The main goal is not an immediately translatable treatment, but a solid foundation from which further research and treatments can be built upon.

Materials and Methods

Bacterial Expression of SBP tagged MYC

SBP-MYC was purified from E. coli bacteria previously transformed by Dr. Caballero with a MYC vector pET28-6xHis-SBP-TEV (Addgene). A frozen glycerol stock of this bacteria was thawed and incubated overnight at 37° C using 25 mg/mL KanaMYCin (ThermoFisher) and 25 mg/mL Chloramphenicol (ThermoFisher) in Lysogeny Broth media (Boston Bio Products). The next day a sample of this was incubated in a 500 mL culture of media at the same temperature and same antibiotic concentration. When the OD600 of the bacteria was measured to be between 0.5-1, and thus in the exponential growth stage, bacterial growth was slowed on ice and expression of MYC with a streptavidin binding peptide (SBP) and Tobacco Etch Virus (TEV) tag attached was induced using 0.5 mM IPTG. After the bacteria culture was incubated overnight at room temperature the culture was centrifuged for 30 minutes at 3000 rpm at 4° C. The pelleted bacteria was resuspended in 20 mL of lysis buffer, a solution composed of 20 mL CelLytic B (Sigma-Aldrich), 0.8 mL Lysozyme (Sigma-Aldrich), 2.6 μL benzonase (Sigma- Aldrich), and 2 tablets of protease inhibitor cocktail (Roche). The lysate was incubated for 30 minutes at 37° C before being centrifuged for 15

minutes at 12,000 g at 4° C, discarding the pellet and storing the supernatant at -80° C for use in both purification and co- immunoprecipitation experiments.

Purification of SBP tagged MYC

300 µL of streptavidin agarose beads (ThermoScientific) were washed twice with cold PBS, centrifuging for 4 minutes at 200 g at room temperature. The streptavidin agarose beads were incubated overnight at 4° C with 10 mL of the previously collected lysate. After this incubation the flow through was stored, henceforth referred to as FT-Beads. The beads remaining in the tube were then washed twice again with cold PBS. To separate the MYC from the SBP-TEV tag and have pure MYC, AcTEV Protease (Invitrogem) was used to bind the TEV part and then cleave MYC from the SBP-TEV tag. The 300 µL of streptavidin agarose beads were incubated overnight at 4° C with a Reaction Mix solution composed of 30 µL of 20X TEV buffer, 6 µL of 0.1 M DTT, 4 µL of AcTEV Protease, and 600 µL of pure water. After incubation with the TEV Reaction Mix, the streptavidin agarose beads were centrifuged for 5 minutes at 200 g at 4° C, and the supernatant was then separately stored, henceforth referred to as FT-TEV. Finally, 300 µL of 2x Laemmli sample running buffer (Bio-Rad) was added to the beads and the tubes were boiled for ten minutes at 90° C, and centrifuged for 5 minutes at 12000 g at room temperature.

20 µL of the Laemmli buffer was added to 20 µL samples of the FT-Beads and FT-TEV and these were boiled for ten minutes at 90° C, centrifuged for 5 minutes at 12000 g at room temperature. These samples, in combination with a 40 µL sample of the boiled beads mentioned above, were run on 4-15% polyacrylamide gels (Bio-Rad) in 1X Tris/glycine/SDS buffer at 200 V for about 45 minutes. One gel was analyzed by Coomassie stain, and the other gel was transferred to a PVDF membrane (Bio-Rad) and analyzed by western blot using a 1:500 dilution of anti-MYC (9E10) mouse monoclonal antibody (Santa Cruz Biotech) and a 1:2000 dilution of anti-mouse monoclonal antibody (Santa Cruz Biotech). Blots were developed using a 1:1 solution of the two West Pico Stablizer/Enhancer solutions (ThermoScientific) and visualized through chemiluminescence using the ChemiDoc MP Imaging System machine (Bio-Rad).

Co-immunoprecipitation of SBP-MYC and MAX with 10074-G5

Co-immunoprecipitation assays were performed to look at the ability of SBP-MYC to bind to and pull down MAX in vitro while under the condition of being incubated with modulators of this interaction. This first set of assays use the positive MYC/MAX heterodimer inhibitors 10074- G5 (Selleck Chem) or PY9 (Sigma-Aldrich) as the compounds modulating the interaction between SBP-MYC and MAX. A binding buffer solution was produced that was composed of pure water, 10x Binding buffer (EMSA kit, ThermoScientific) diluted to a 1x final concentration, 50% glycerol diluted to a 2.5% final concentration, 100 mM MgCl2 diluted to a 5 mM final concentration, 1% NP-40 diluted to a 0.05% final concentration. This solution is henceforth referred to as simply binding buffer. 1 mL of Streptavidin agarose beads (ThermoScientific) were washed twice with cold PBS, centrifuging for 5 minutes at 200 g at 4° C. 100 µL of these washed beads were aliquoted into 4 tubes to be incubated with 100 µL of the SBP-MYC lysate previously frozen and 800 µL of the binding buffer. Another tube was prepared to be the negative control, containing only the beads and the binding buffer, no SBP-MYC. All 5 of these were incubated for 1.5 hours at room temperature and then centrifuged for 3 minutes at 300 g at 4° C, removing the supernatant. To all of these tubes 1 µL of pure His-tagged MAX protein (Abcam) and 889 µL of the binding buffer was added. The positive control tubes additionally had increasing amount of MYC/MAX heterodimer inhibitors 10074-G5 (25 µM, 50 µM, and 100 µM) or PY9 (10 µM, 15 µM, 20 µM) added to the solution, whereas the negative control and test tubes just had a final concentration of 1% DMSO, without additional drug. Again all 5 of these were incubated for 1.5 hours at room temperature and then centrifuged for 3 minutes at 300 g at 4° C, removing the supernatant. For the last time, the streptavidin agarose beads were washed twice with cold PBS.

100 µL of the Laemmli buffer was added to 100 µL samples of the beads and boiled for ten minutes at 90° C, centrifuged for 5 minutes at 12000 g at room temperature. 30 µL of each sample were run on 4-15% polyacrylamide gels (Bio-Rad) in 1X Tris/glycine/SDS buffer at 200 V for about 45 minutes. The gel was transferred to a PVDF membrane (Bio-Rad) and analyzed by western blot using a 1:500 dilution of anti-MYC (9E10) mouse monoclonal antibody (Santa Cruz Biotech), a 1:500 dilution of anti-MAX (C-124) rabbit monoclonal antibody (Santa Cruz Biotech), and 1:2000 dilutions of anti-mouse and anti-rabbit monoclonal antibodies (Santa Cruz Biotech). Blots were developed using a 1:1 solution of the two West Pico Stabilizer/Enhancer solutions (ThermoScientific) and visualized through chemiluminescence using the ChemiDoc MP Imaging System machine (Bio-Rad).

Co-immunoprecipitation of MYC and MAX from P493-6 cell lysates with KI-MS2-008

The protocols and experiments described in this section were all performed by my mentor, graduate student Andrew Chen. P493-6, a Burkitt's lymphoma cell line which is engineered to express high levels of MYC in the absence of doxycycline, was cultured. 400 million P493-6 cells, as counted by the Cellometer Auto T4 Bright Field Cell Counter (Nexcelom Bioscience), were pooled and washed twice in cold PBS. Cells were resuspended in 10 mL of modified RIPA lysis buffer with 0.1% sodium deoxycholate, 1% NP-40, 0.004% buffer benzonase, 200 mM NaCl, 50 mM Tris, 1X protease inhibitor (Sigma-Aldrich), and 1X phosphatase inhibitor (Roche). After incubating on ice for 15 minutes, the cells were centrifuged for 10 minutes at 14000 g at 4° C.

Following this a Bradford assay was performed to determine the total protein concentration of the cell lysate (the supernatant from the previous step), and 4 mg of total protein was incubated for thirty minutes with either 2 µL DMSO or 20 µM KI-MS2-008 and topped off to a total volume of 1 mL with Ab Binding and Washing Buffer from the magnetic Dynabeads kit (ThermoScientific). Then the protein/compound mixtures were incubated overnight at 4° C with either 20 µL MAX (C-17) rabbit antibody (Santa Cruz Biotech) or 10 µL of normal rabbit IgG (Santa Cruz Biotech). 50 µL of the magnetic Dynabeads per tube were placed on a magnetic stand for a minute to settle the beads, and washed in 200 µL of the Ab Binding and Washing Buffer before transferring the protein/compound/antibody mixture prepared the night prior into the tubes. These new tubes were incubated for one hour at 4° C and then placed on the

magnetic stand again to settle the beads, saving the flow through to run on the gel. This was washed three times using 200 µL of the Washing Buffer, saving the washes for the gel as well.

30 µL of 2X SDS buffer was added to Dynabead tubes and boiled for five minutes at 90° C, using the magnetic stand to settle the beads and separate the elution. 20 µL of each sample were run on 4-15% polyacrylamide gels (Bio-Rad) in 1X Tris/glycine/SDS buffer at 100 V for 10 minutes, followed by 200 V for about 35 minutes. The gel was transferred to a PVDF membrane (Bio-Rad) and analyzed by western blot using a 1:500 dilution of anti-MYC (9E10) mouse monoclonal antibody (Santa Cruz Biotech), a 1:500 dilution of anti-MAX (H-2) mouse monoclonal antibody (Santa Cruz Biotech), and 1:2000 dilutions of anti-mouse monoclonal antibodies (Santa Cruz Biotech). Blots were developed using a 1:1 solution of the two West Pico Stabilizer/Enhancer solutions (ThermoScientific) and visualized through chemiluminescence using the ChemiDoc MP Imaging System machine (Bio-Rad).

Cell Viability Assays

All cell lines were obtained from the MIT Koch Institute's High Throughput Screening facility. The suspension human breast cancer cell line HCC1599 was cultured in Roswell Park Memorial Institute (RPMI) 1640 Medium (ThermoFisher), the adherent human breast cancer cell line MDA-MB-468 was cultured in Dulbecco's Modified Eagle's Medium (DMEM) (ThermoFisher), and the adherent human brain cancer cell line U87MG was cultured in Eagle's Minimum Essential Medium (EMEM) (ATCC). All culture media was prepared with 10% Fetal Bovine Serum (FBS) (ATCC) and 1% Penicillin/StreptoMYCin (P/S) (Corning Incorporated).

Cell cultures were grown using an incubator kept at 37° C at 5% CO2 and cultures were split every 3 to 5 days, as recommended according to the manufacturer's instructions. The compounds used are BET inhibitor JQ1 (Sigma-Aldrich) and PI3K inhibitor GDC-0941 (Selleck Chem), and KI-MS1-001 and KI-MS2-008 which are being produced internally at the Koehler Lab. All four were serially diluted two-fold in DMSO, from 40 µM to 0.156 µM and later from 40 µM to 0.00122 µM, while being kept at a final concentration of 0.4% DMSO.

50 µL of cells were plated in sterile 96 well plates (Corning Incorporated) at a density of 5,000 cells per well, two plates per condition. Suspension cells were directly treated with 50 µL of either JQ1, GDC-0941, KI-MS1-001, KI-MS2-008 or DMSO and placed into the incubator, while adherent cells were first incubated overnight before replacing the culture media and then treating them with compound or DMSO. Each cell line was incubated in compounds or DMSO for both 3 days and 5 days, time points after which a plate would be visualized. 100 µL of CellTiter-Glo reagent (Promega) was added to each well, and plates were incubated for 10 minutes at room temperature on a platform shaker. Luminescence was recorded using a Tecan Infinite 200 Pro spectrophotometer using an integration time of 0.5 second/well. Luminescence values were graphed, and IC50 values calculated, using Prism software (Graphpad).

Following testing of cell viability under compound and DMSO conditions, cells were also prepared for western blotting against MYC, p-AKT, and GAPDH. Using sterile 12 well plates (Greiner Bio-One) for the HCC1599 suspension cells and sterile 6 well plates (VWR) for the MDA-MB-468 adherent cells, 3 million cells from each cell line were incubated for 24 hours with two-fold serial dilutions, from 10 µM to 0.156 µM, of the four compounds at 0.4% DMSO. Suspension cells were directly treated with the compounds while adherent cells were first incubated overnight in culture media before being treated with compounds. Cells were pelleted and washed with cold PBS, centrifuging for 5 minutes at 200 g at 4° C. Then cells were resuspended in 60 µL of modified RIPA lysis buffer with 0.1% sodium deoxycholate, 1% NP-40, 0.004% buffer benzonase, 200 mM NaCl, 50 mM Tris, 1X protease inhibitor (Sigma-Aldrich), and 1X phosphatase inhibitor (Roche). Total protein concentration was determined using a BCA assay, and samples meant to run on the gel were either prepared to have a consistent 30 µg concentration of protein or 50 µg, depending on the total protein concentration from the cell line.

Enough 2X SDS buffer was added to samples for a 1:1 ratio of sample and running buffer, and samples were boiled for five minutes at 90° C, centrifuged for 5 minutes at 12000 g at room temperature. Samples were run on 4-15% polyacrylamide gels (Bio-Rad) in 1X Tris/glycine/SDS buffer at 200 V for about 45 minutes. The gel was transferred to a PVDF membrane (Bio-Rad) and analyzed by western blot using a 1:500 dilution of anti-MYC (9E10) mouse monoclonal antibody (Santa Cruz Biotech), a 1:500 dilution of anti-MAX (C-124) rabbit monoclonal antibody (Santa Cruz Biotech), a 1:1000 dilution of anti-p-AKT (T308) rabbit monoclonal antibody (Cell Signaling), a 1:1000 dilution of anti-GAPDH (14C10) rabbit monoclonal antibody (Cell Signaling), and 1:2000 dilutions of anti-mouse and anti-rabbit monoclonal antibodies (Santa Cruz Biotech). Blots were developed using a 1:1 solution of the two West Pico Stabilizer/Enhancer solutions (ThermoScientific) and visualized through chemiluminescence using the ChemiDoc MP Imaging System machine (Bio-Rad).

Supplementary Figures

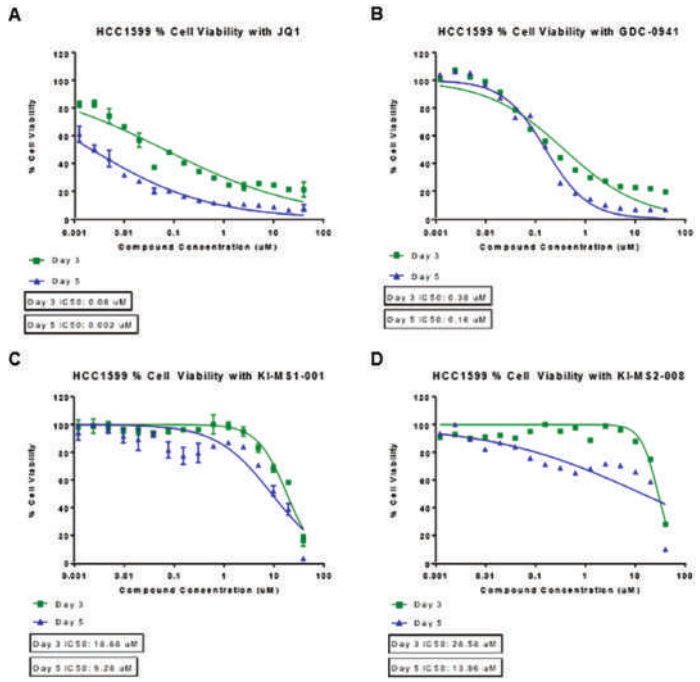


Figure S4: HCC1599 cell viability assay data, third biological replicate. Like the second replicate, all drug compounds again displayed lower IC50 values on Day 5 in comparison to Day 3. (A) All of the drug compounds seemed to be capable of reducing cell viability, with JQ1 to being the most potent out of all four, going so low as 0.002 uM. (B) GDC-0941 also had IC50s less than 1 uM. (C) KI-MS1-001 had higher IC50, around 9 uM. (D) This time KI-MS2-008 had the highest IC50 of around 14 uM.

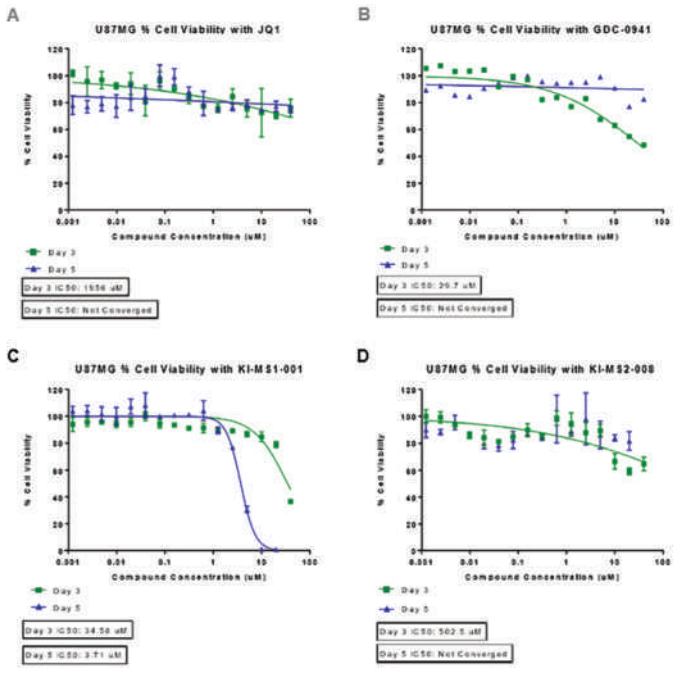


Figure S5: U87MG cell viability assay data, third biological replicate. Like the second replicate, none of the drug compounds seemed effective. (A, B, D) The Day 5 IC50 values for JQ1, GDC-0941, and KI-MS2-008 could not be determined because the graphs did not converge. (C) The IC50 values for KI- MS1-001 went down from Day 3 to a Day 5 IC50 value of 3.7 uM, but since this was not the case for any of the other drugs, this is probably due to error.

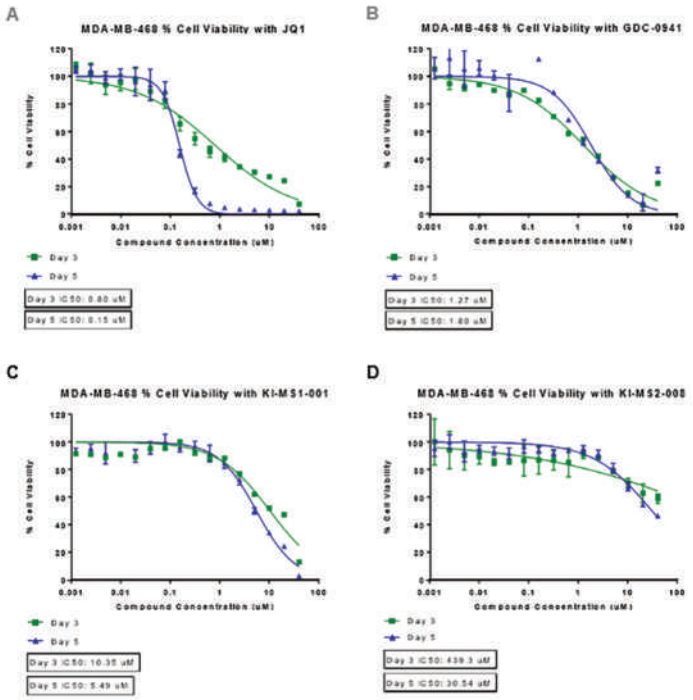


Figure S6: MDA-MB-468 cell viability assay data, third biological replicate. Like the second replicate, all drug compounds again displayed lower IC50 values on Day 5 in comparison to Day 3. (A) Like the second replicate, JQ1 and GDC-0941 had IC50s of around 1 uM, with JQ1 seeming to be the most potent out of all four drug compounds, having an IC50 of 0.15 uM. (B) GDC- 0941 was a bit higher, with an IC50 of 1.8 uM. (C) However, unlike in the second replicate, KI-MS1-001 had a higher IC50, around 5.5 uM. (D) Most unexpectedly, KI-MS2-008 was also different here from the second replicate, having the highest IC50 of 30.5 uM.

References

Adhikary, S., & Eilers, M. (2005). Transcriptional regulation and transformation by MYC proteins. Nat Rev Mol Cell Biol, 6(8), 635-645.

Arvanitis, C., & Felsher, D. W. (2006). Conditional transgenic models define how MYC initiates and maintains tumorigenesis. Semin Cancer Biol, 16(4), 313-317. doi:10.1016/j.semcancer.2006.07.012

Boxer, L. M. D., Chi. V. (2001). Translocations involving MYC and MYC function. Oncogene, 20(40), 5595-5610. doi:10.1038/sj.onc.1204595

Bradner, J. E., McPherson, O. M., & Koehler, A. N. (2006). A method for the covalent capture and screening of diverse small molecules in a microarray format. Nat Protoc, 1(5), 2344- 2352. doi:10.1038/nprot.2006.282

Clemons, P. A., Bodycombe, N. E., Carrinski, H. A., Wilson, J. A., Shamji, A. F., Wagner, B. K., . . . Schreiber, S. L. (2010). Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles. Proc Natl Acad Sci U S A, 107(44), 18787-18792. doi:10.1073/pnas.1012741107

Dang, C. V. (1999). MYC Target Genes Involved in Cell Growth, Apoptosis, and Metabolism. Molecular and Cellular Biology, 19(1), 1-11.

Dang, C. V. (2012). MYC on the Path to Cancer. Cell, 149(1), 22-35. doi:10.1016/j.cell.2012.03.003

Darnell, J. E., Jr. (2002). Transcription factors as targets for cancer therapy. Nat Rev Cancer, 2(10), 740-749. doi:10.1038/nrc906

Delmore, J. E., Issa, G. C., Lemieux, M. E., Rahl, P. B., Shi, J., Jacobs, H. M., . . . Mitsiades, C.S. (2011). BET bromodomain inhibition as a therapeutic strategy to target MYC. Cell, 146(6), 904-917. doi:10.1016/j.cell.2011.08.017

Felsher, D. W. (2010). MYC Inactivation Elicits Oncogene Addiction through Both Tumor Cell– Intrinsic and Host-Dependent Mechanisms. Genes & Cancer, 1(6), 597-604. doi:10.1177/1947601910377798

Jain, M., Arvanitis, C., Chu, K., Dewey, W., Leonhardt, E., Trinh, M., . . . Felsher, D. W. (2002). Sustained Loss of a Neoplastic Phenotype by Brief Inactivation of MYC. Science, 297(5578), 102-104. doi:10.1126/science.1071489

Li, Z., Van Calcar, S., Qu, C., Cavenee, W. K., Zhang, M. Q., & Ren, B. (2003). A global transcriptional regulatory role for MYC in Burkitt's lymphoma cells. Proc Natl Acad Sci U S A, 100(14), 8164-8169. doi:10.1073/pnas.1332764100

Liu, P., Cheng, H., Santiago, S., Raeder, M., Zhang, F., Isabella, A., . . . Zhao, J. J. (2011). Oncogenic PIK3CA-driven mammary tumors frequently recur via PI3K pathway- dependent and -independent mechanisms. Nature medicine, 17(9), 1116-1120. doi:10.1038/nm.2402

Muellner, M. K., Uras, I. Z., Gapp, B. V., Kerzendorfer, C., Smida, M., Lechtermann, H., . . . Nijman, S. M. (2011). A chemical-genetic screen reveals a mechanism of resistance to PI3K inhibitors in cancer. Nat Chem Biol, 7(11), 787-793. doi:10.1038/nchembio.695

Orian, A., van Steensel, B., Delrow, J., Bussemaker, H. J., Li, L., Sawado, T., . . . Eisenman, R.N. (2003). Genomic binding by the Drosophila MYC, Max, Mad/Mnt transcription factor network. Genes & Development, 17(9), 1101-1114. doi:10.1101/gad.1066903

Pop, M. S., Stransky, N., Garvie, C. W., Theurillat, J.-P., Lewis, T. A., Zhong, C., . . . Garraway, L. A. (2014). A small molecule that binds and inhibits the ETV1 transcription factor oncoprotein. Molecular cancer therapeutics, 13(6), 1492-1502. doi:10.1158/1535- 7163.MCT-13-0689

Prochownik, E. V., & Vogt, P. K. (2010). Therapeutic Targeting of MYC. Genes Cancer, 1(6), 650-659. doi:10.1177/1947601910377494

Riss, T. L., Moravec, R. A., & Niles, A. L. (2011). Cytotoxicity Testing: Measuring Viable Cells, Dead Cells, and Detecting Mechanism of Cell Death. In M. J. Stoddart (Ed.), Mammalian Cell Viability: Methods and Protocols (pp. 103-114). Totowa, NJ: Humana Press.

Soucek, L., Whitfield, J., Martins, C. P., Finch, A. J., Murphy, D. J., Sodir, N. M., . . . Evan, G. I. (2008). Modelling MYC inhibition as a cancer therapy. Nature, 455(7213), 679-683. doi:10.1038/nature07260

Stratikopoulos, E. E., Dendy, M., Szabolcs, M., Khaykin, A. J., Lefebvre, C., Zhou, M. M., & Parsons, R. (2015). Kinase and BET Inhibitors Together Clamp Inhibition of PI3K Signaling and Overcome Resistance to Therapy. Cancer Cell, 27(6), 837-851. doi:10.1016/j.ccell.2015.05.006

van Riggelen, J., Yetil, A., & Felsher, D. W. (2010). MYC as a regulator of ribosome biogenesis and protein synthesis. Nat Rev Cancer, 10(4), 301-309.

Vita, M., & Henriksson, M. (2006). The MYC oncoprotein as a therapeutic target for human cancer. Seminars in Cancer Biology, 16(4), 318-330. doi:http://dx.doi.org/10.1016/j.semcancer.2006.07.015

Wang, H., Hammoudeh, D. I., Follis, A. V., Reese, B. E., Lazo, J. S., Metallo, S. J., & Prochownik, E. V. (2007). Improved low molecular weight MYC-Max inhibitors. Molecular cancer therapeutics, 6(9), 2399.

Yin, X., Giap, C., Lazo, J. S., & Prochownik, E. V. (2003). Low molecular weight inhibitors of MYC-Max interaction and function. Oncogene, 22(40), 6151-6159.

Understanding Immigration: A Closer Look at the Fiscal Impact of Immigrant Characteristics

Jean Thirouin '17

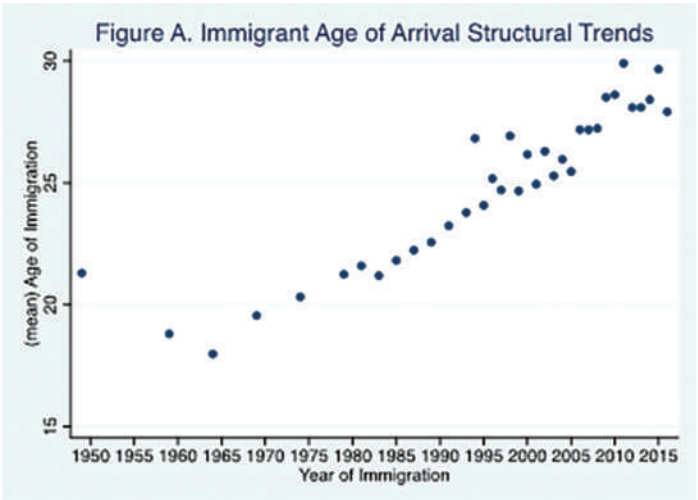
This paper attempts to further research on the fiscal impact of immigration. My findings are illustrated by computing the net fiscal impact, in present value terms, of admitting one additional immigrant, conditional on education, gender, and age at the time of immigration. I demonstrate that the average immigrant arriving past age 34 has a lifetime negative fiscal impact. Additionally, a college educated immigrant arriving past age 52 will have a lifetime negative fiscal impact while a non-college educated immigrant will roughly have a lifetime negative fiscal impact, regardless of age at arrival. Further, I confirm that age at arrival matters, and determine that arrival prior to working age influences educational attainment. Finally, I provide a household life-cycle model that sheds light on the fiscal contribution of immigrating families. All immigration referred to in this paper is legal.

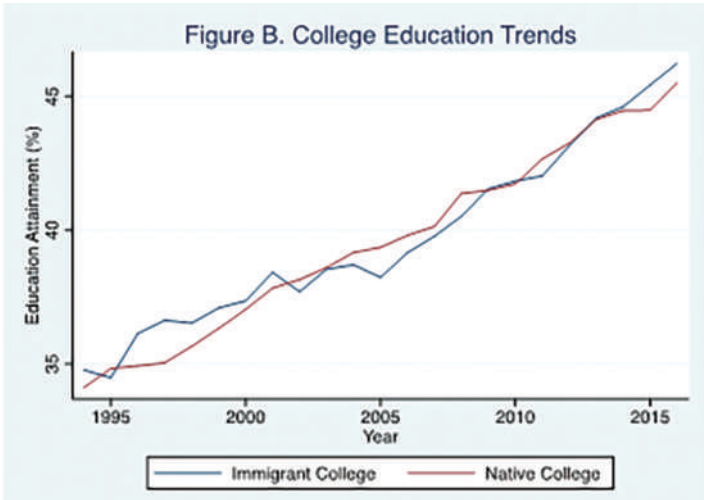
With an aging population and a social security system set to run out of funds in the foreseeable future, is immigration the key to slowing the decline in the working-age share of the population while helping bolster a strained fiscal deficit? This question is at the heart of the public policy debate on immigration. The key to answering this question lies in understanding the age structure of immigrants and their life-cycle fiscal impact.

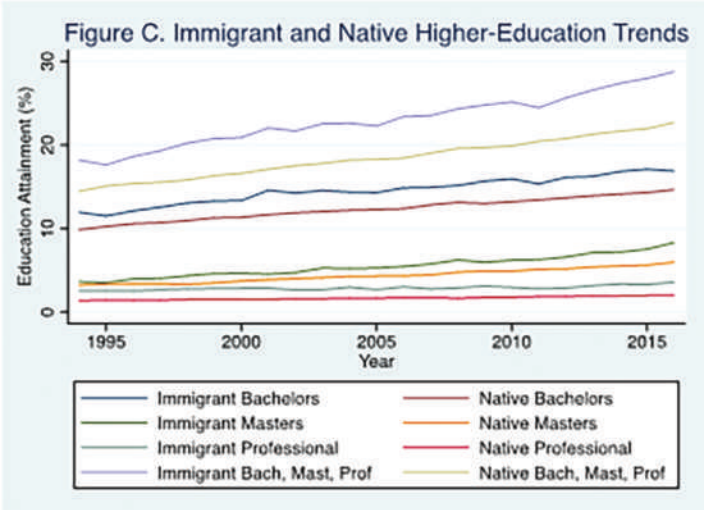
Immigration has greatly increased in the past decades. From 1940-1950 there were about 1mil immigrants, 1950-1960 had 2.5mil immigrants, 1960-1970 had 3.3mil immigrants, 1970-1980 had 4.5mil immigrants, 1980-1990 had 5.7mil immigrants, 1990-2000 had 11.3mil immigrants, and 2000-2010 had 8.8mil immigrants (Gibson 1999). Along with the increase in immigration has been an increase in public opinion on the best public policy pathway. Some have viewed immigrants as a "fiscal drain", while others have believed in their ability to provide a solution to the aging population problem. Skilled workers that would immigrate early and immediately contribute through taxes would likely lead to large positive net fiscal effects, even accounting for the subsequent costs of retirement (Kjetil 2000).

A particular interest of immigration is their impact on the host country's welfare system. The United States' pay-as-you-go system relies on the working age population to support retirees, which lends itself to the argument in favor of increased immigration. From 1980 to 2010, though, the share of inflows of immigrants aged 50 to 74 increased from 8.9% to 14.5% (Yi 2014). As can be seen in Figure A, the mean age of immigrants has been steadily rising since the 1970s, with women, on average, older than men. If immigrants are entering the country at a later age, they may be worsening the aging situation. On the other hand, Figure B and C paint a different picture as they show an upward trend in the educational attainment of immigrants. Between 1950 and 2007, the foreign-born share of employees in the U.S with a masters, professional, or doctorate degree rose from 5.9% to 18.1% (Peri 2010). They may be coming in later, but if they are more skilled, their impact on the welfare system may also be larger.

Although immigration is a hot policy topic, there have only been three major changes in U.S immigration policy in recent history: 1) the 1996 Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) or Welfare Reform, 2) the 1920s national-origin quotas, and 3) the 1965 elimination of such quotas (Klopfenstein 1998). The 1996 Welfare Reform served to restrict







immigrant access to Federal welfare benefits such as Medicare and food stamps for their first five years in the country, however, states could grant aid out of their own funds. Additionally, in 2004 the annual cap for new H-1B visas was lowered from 195,000 to 65,000 – essentially an attempt to reduce competition between similarly-educated immigrants and natives (Peri 2010).

What is the best course of action for the government to solve a growing fiscal deficit tied to an aging population problem? Should the government focus more on changing the level or mix of immigrants (Auberbach Oreopoulos 2000)? The goal of this paper is to shed light on the quantitative aspect of this debate.

The remainder of the paper proceeds as follows. I provide a brief literature review in Section I. Section II describes my data, Sections III, IV, V provide my method along with my results, and Section VI constitutes a household life-cycle model. Section VII discusses limitations, government policy implications, and concludes.

Literature Review

Immigration has a growing body of literature with two mains sides: 1) the labor market outcomes, and 2) the fiscal impact (Borjas 2013). Much of the focus has been on the demand side of the labor market impact caused by immigrants. More specifically, attention has been paid to the effects that immigrants have had on native wages. While research has shown immigrants to lower native wages, especially those with less than a college level education, others have argued for the complementary nature of immigrants (Ottaviano Peri 2011). My paper focuses on the fiscal impact of immigrants therefore I will mainly reference the literature on that front.

Despite the strong implications of immigration for public finance, there is a limited amount of literature addressing the cost-benefit life-cycle aspect of immigration (Friedburg Hunt 1995). However, the related literature is important in formulating hypotheses and making assumptions in my later models. One main focus has been the importance of immigrating at an early age (OECD) (Shaafsma Sweetman 2001)(Myers et al. 2009)(Seeborg Sandford 2003). Comparably, human capital accumulation and language proficiency have been determined to be two of the most important characteristics of immigrants at arrival, closely tying in to age of arrival effects (Lagakos 2016). Intuitively, an immigrant arriving as a young child has a much higher likelihood of assimilating into the culture and surpassing the language barrier than an immigrant in his late thirties (Hoyt Chin 2016). Additionally, immigrants from poor countries will tend to accumulate much less human capital in their birth countries before migrating (Lagakos 2016).

No clear consensus has been reached about the use of welfare by immigrants, however. While some studies have shown that longer time spent in the host country (Sweden) have led to decreased rates of welfare use, others have shown the exact opposite (Hansen Lofstrom 2010). Another issue concerning welfare literature has been the lack of separation between welfare usage and welfare eligibility (Pekkala Kerr 2011). This is especially important to consider due to changes in eligibility over time.

There are two main techniques for evaluating the fiscal impact of immigrants. The first estimates the growth to GDP due to growth in labor supply (Borjas 1995)(Freeman 2006)(Drinkwater et al. 2007). The second relies mostly on accounting methods and estimating the total cost and benefits that immigrants will have on the economy, which varies greatly by stage of life (Pekkala Kerr 2011). The accounting method is the type my paper focuses on. More specifically, my paper sheds light on the varying net impact at all stages in an immigrant's life. As well, I expand on my different life-cycle models to estimate household impacts of migrating families for alternate family structures.

Data

My data comes from the March Annual Social and Economic Supplement (ASEC) supplemental survey of the Current Population Survey (CPS). It is all downloaded from the Integrated Public Use Microdata Series (IPUMS) website which allows the identical coding of variables across time for easier cross-time comparison. Additionally, IPUMS-CPS provides all estimations with the person weights in the census data files (Flood King 2015). My sample consists of pooled microdata, information on individual persons and households, for the years 1994 to 2016. Data from previous years was not used since the age classification of immigrants could only be constructed from a variable in my time frame.

The IPUMS-CPS data source provides variables crucial to my research such as age, year of immigration, gender, education, taxes, benefits received, among others. I classify immigrants by marking immigration population (impop) = 1 if they answered the question about year of arrival. Unlike most of the variables in the CPS, the major tax variables such as FEDTAX, STATETAX, and FICA are not the result of direct questioning of respondents. Instead, the values for the variables come from the Census Bureau's tax model which was updated in the fall of 2004 to produce more accurate tax estimates. The model simulates tax returns for each individual to produce the estimates needed by incorporating information from non-CPS sources such as the Internal Revenue Service's Statistics of Income series, the American Housing Survey, and the State Tax Handbook. I adjust all monetary values for inflation to the 1999 CPI indicator using a predefined IPUMS-CPS variable which multiplies a factor for each non-1999 year to a dollar amount in that year.

Net Yearly Contribution Model

Net Yearly Contribution - Average

The most important variable I constructed for each observation in my data set is an individual's net contribution (netcontri) to the

government. This is defined as their total taxes paid (taxtot) minus the total benefits received from the government (incgov). Since my data set did not contain predefined variables for total taxes nor total benefits received, using IPUMS' recommendations, I constructed them by summing across individually reported values of its components. The compositions are defined below:

Since netcontri is central to the remainder of my analysis, it is

$$\textbf{Tax Total} = \textit{Federal Tax} + \textit{State Tax} + \textit{FICA}$$

$$\textbf{Total Benefits} = \textit{Unemployment Income} + \textit{Workers Compensation} + \textit{Veteran's Association Benefits} + \textit{Social Security} + \textit{Welfare Income}$$

crucial that the variables it is comprised of are as closely reflective to the actual values as possible. I perform a sensitivity analysis in order to determine their viability. Taxtot has relatively little room for error since it already accounts for the three major sources of tax revenue – federal tax, state tax, and FICA. However, an individual's income received from the government is more complex since it can come from many different sources. I experiment with alternate compositions such as the inclusion of disability income, survivor's benefits, and earned income tax credit. Plotting the average difference between these compositions and the one mentioned above by age, I notice a negligible impact when looking at immigrants only. I run the same analysis for natives and notice a slightly greater impact past the age of 50 yet it remains insignificant. I plot both immigrants and natives separately because I am (correctly) predicting that the composition of variables such as survivor's benefit differs for immigrants and natives.

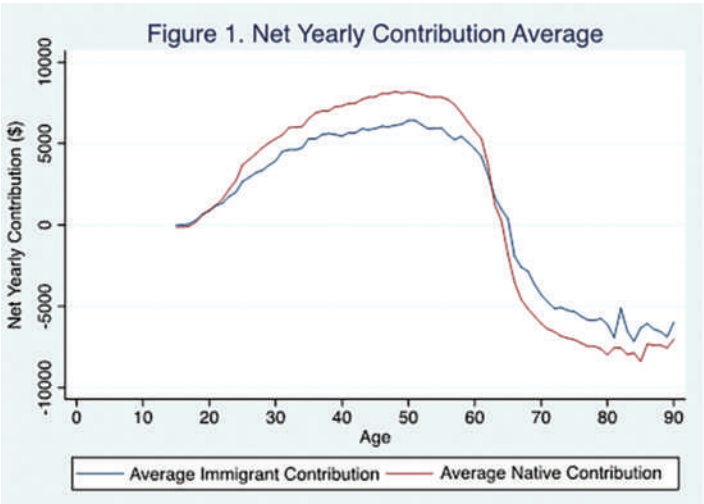
After constructing my net contribution variables for each observation in my data, the first step in my analysis is to find the average contribution at each age. Plotting the results gives me a first pass look at the impact an immigrant has on the government budget at each age in their life. As can be seen in Figure 1, an immigrant's impact over their life follows an intuitive understanding of one's life-cycle earnings. Earnings begin around 15 years of age and sees its greatest growth between the early twenties to the mid-thirties, where it stands around $5,300/year. It begins roughly leveling off to around $6,200/year, the peak earnings stage during the fifties. From there, the decline begins, which becomes sharp around the early sixties, the age of retirement, and roughly levels off around -$6,000/year for the remainder of an immigrant's life.

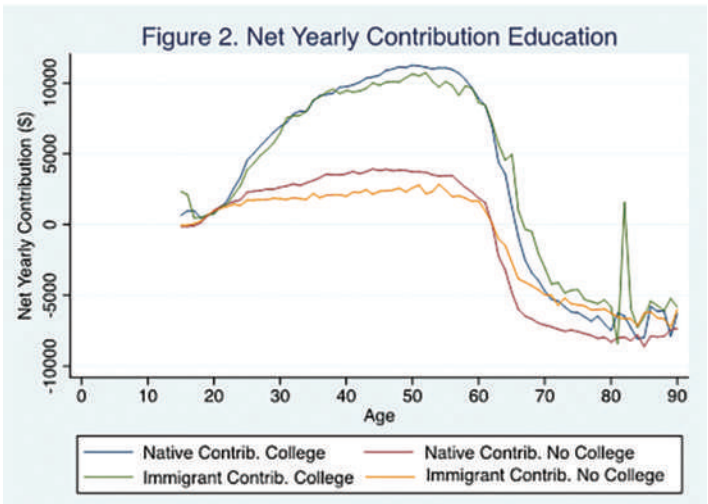
To compare immigrants' average earnings to natives, I run the same analysis restricting the data to natives only. Shown in Figure 1, a native's average impact at each age follows a similar pattern with slight differences. Earnings begins at age 15 and rises faster than immigrants until the mid-thirties where it reaches around $6,500/year. Instead of leveling off, though, earnings continue to increase, albeit at a slower rate, until the fifties where they peak at $8,200/year. A similar sharp decline can be seen around the early sixties before roughly leveling off at -$7,500/year for the remainder of their life. Figure 1 shows the aggregation of all immigrants and all natives, whether male, female, college educated, living in the US for a short/long time, etc. In other words, the results discussed above can be generalized to "an average immigrant" and "an average native". However, when facing the decision to accept a foreigner into the country, other differentiating characteristics mentioned above are observed. Below I provide a contribution analysis based on the main characteristics observed when a foreigner attempts to immigrate to the United States, limited to the ones observed in my data set.

Net Yearly Contribution - Education

I continue my analysis of an immigrant's yearly impact on the government budget by taking into account their level of education. I construct a binary variable college that is defined as 1 if an individual has completed at least a year of college all the way up to doctorate degrees, and 0 for any educational attainment below one year of college. I then find the average contribution per year at each age for both immigrants and natives, with and without a college education.

The results for college and non-college educated immigrants and natives are displayed in Figure 2. We can see that both college and non-college educated immigrants follow a very similar life-cycle pattern mentioned earlier in our "average immigrant" case. The stark difference is revealed in their earnings growth and peak earnings. An immigrant that has at least one year of college education will, on average, peak at around $10,000/year net contribution in his mid-fifties. A less educated immigrant, on average, will peak and stagnate closer to $2,000/year. Said in another way, a college educated immigrant, on average, will have a five times greater positive





impact on the government budget than a non-educated immigrant.

Looking at the impact that education can have on an immigrant's earnings clearly reveals its importance, yet even more can be learned when compared to the impact that education has on a native's earnings. A college educated native mirrors the earnings of a college educated immigrant, with peak earnings slightly surpassing $11,000/year. Interestingly, non-college educated natives perform much better than non-college educated immigrants. A non-educated native reaches peak contribution closer to $3,500/year, or almost twice that of an identical immigrant.

Figure 2 reveals another comparison between immigrants and natives. We saw that natives are reaching higher peak earnings during their working life and therefore have a higher net contribution. Once retirement age in the early sixties is attained, though, we now see natives have a much higher negative contribution than immigrants. Are contributions during one's working life offsetting the income received from the government post-retirement? If not, at what age is the cutoff? This phenomenon motivates the life-cycle model I build later in the paper that can shed light on immigrant and native impacts on the government budget throughout the entirety of their life.

It is impossible to tell strictly from the graphs what factors may be causing the discrepancy in earnings between immigrants and natives. Non-educated immigrants most likely suffer from the language barrier, strongly decreasing their already limited employment options. A non-educated native may be able to benefit from relatively higher earnings based on an ability to be employed in communication-based jobs. At higher levels of education, however, immigrants don't suffer as much from the language barrier as they can go into high earnings fields that are focused more on quantitative skills rather than communicative ones.

Net Yearly Contribution - Gender

Should the government have the right to discriminate against incoming immigrants depending on their education? If so, should they be able to do the same regarding an immigrant's gender? Whether or not they should, a lot can be learned by analyzing the differential earnings of men and women. The trends observed earlier in both "average immigrants" and educated immigrants can further be explained by separating the data according to an immigrant's gender. Additionally, these trends prove useful later in understanding the dynamics of the household life-cycle model.

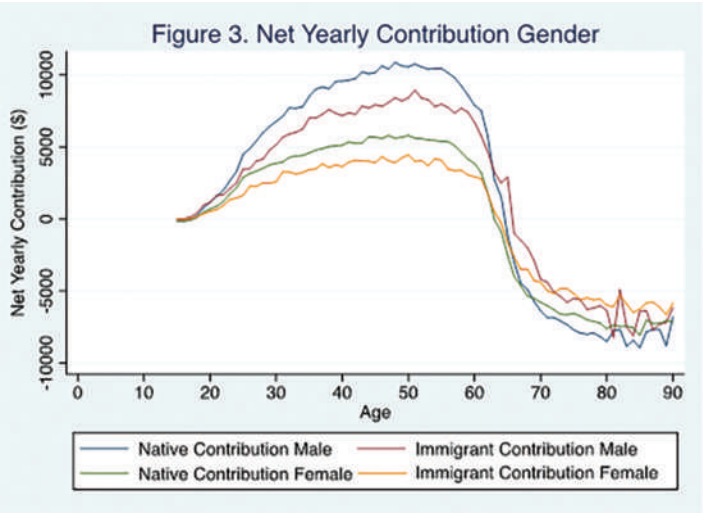


Figure 3 illustrates the net contribution gap between male and female immigrants. It is evident that men have much higher earnings growth throughout their working life. Both men and women seem to reach peak contributions around their fifties, but men on average contribute close to $8,500/year while women peak closer to $4,000/year. Once retirement is reached, both men and women follow similar net contribution trajectories into the negatives.

Figure 3 also allows for the comparison of the effect of gender on the net contributions of both natives and immigrants. Native men and women, similarly to immigrants, have differential impacts on the government budget at each age before retirement. However, the men's highest contributions are about $10,500/year with the women's around $5,500/year, both almost $2,000 more than immigrants for each gender. Unlike immigrants, native women have a slightly smaller negative impact than native men into their seventies and beyond.

General Lifecycle Model

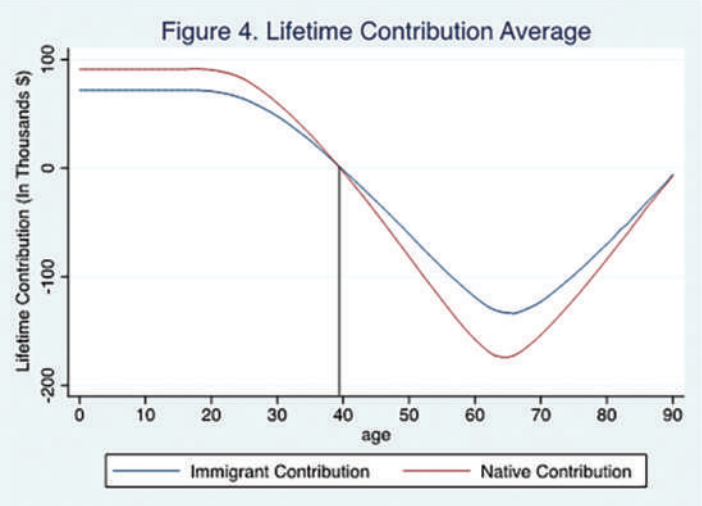
The net contribution model reveals the general trends that are typically seen in an immigrant's life-cycle impact on the government budget at each age. Throughout their working life, fifteen to early sixties, they have a net positive effect on the budget each year by paying more in taxes than they are receiving in welfare. Additionally, the model provides insight into the effect of differing characteristics on the magnitude of immigrants' impacts. As shown, education and gender play an important role in predicting future earnings.

However, the model fails on two fronts: 1) it doesn't reveal whether an immigrant (or native) has a net positive impact on the government budget over the entirety of their life, and 2) it does not factor in a potentially crucial factor, the amount of time spent in the United States. I build two different models to address each problem. The first sums an immigrant's contribution over their lifetime, showing at what age an incoming immigrant no longer benefits the country's government budget. The second expands on this model by restricting the sample for each age of arrival, essentially allowing me to elicit the effect that duration in the country has on earnings.

Lifecycle Model 1 - Average

My first pass life-cycle model is generalized to the "average immigrant" and builds off the previous yearly net contribution model. I use the net contribution at each age and sort them negatively, from oldest to youngest. I then sum up every yearly contribution. Each resulting data point, plotted in Figure 4, indicates an immigrant's future impact on the government budget at that age. This allows me to answer the question, "What is an immigrant's lifetime impact if they arrive at age x?".

Figure 4 also indicates the point at which positive contributions during the working life offset negative contributions after retirement. An immigrant arriving at age 39 will break even in terms of lifetime contributions to the government budget. From a pure government budget standpoint, this indicates that any immigrant over the age of 39 should not be allowed to enter the United States. In Figure 4, we can also observe the lifetime contributions of natives. It is quite interesting to see that natives have the exact same cutoff point, 39 years of age - even though they contribute



more during their working life, they receive more benefits post-retirement. Lastly, we can see that, on average, an immigrant and native who start off at the same point, age 15, will differ in their lifetime impact by $19,000, with natives contributing $91,000 and immigrants contributing $72,000.

Lifecycle Model 1 - Education

Analogous to the yearly contribution education model, I use the lifetime contribution method developed in the "average immigrant" case to demonstrate the effects of education on an immigrant's total impact to the government budget. Using the previously constructed college variable, I separate the data into college educated and less than one year of college education. Sorting the ages from oldest to youngest, I sum up all observations and plot the resulting data points shown in Figure 5. I include the equivalent analysis for natives to use as a comparison "control" group.

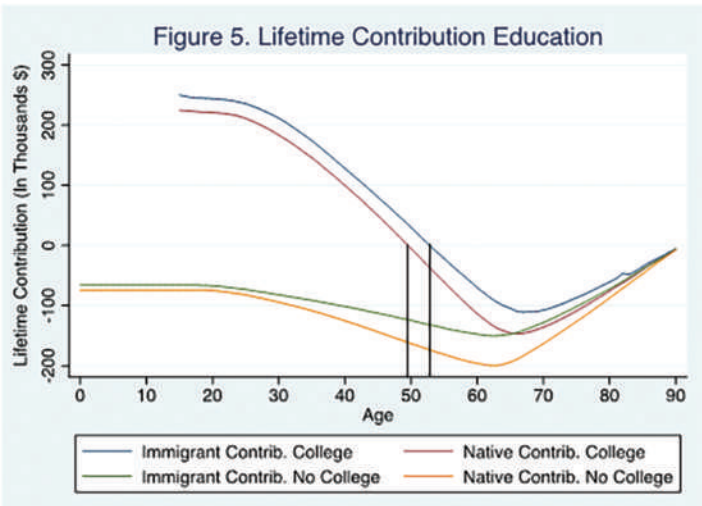
The results seen in Figure 5 illustrate the drastic impact of education that was not as evidently clear from the yearly contribution model. No matter at what age they arrive to the United States, an immigrant with a high school diploma and less will never have a net positive impact on the government budget. In other words, they will always take out more from the United States than they will be able to give back over their entire lifetime. In stark contrast, an immigrant with at least one year of college education and above will be profitable over their entire lifetime so long as they arrive before age 53. In addition, a college educated immigrant who starts working at age 15 will, on average, contribute $250,000 to the government budget.

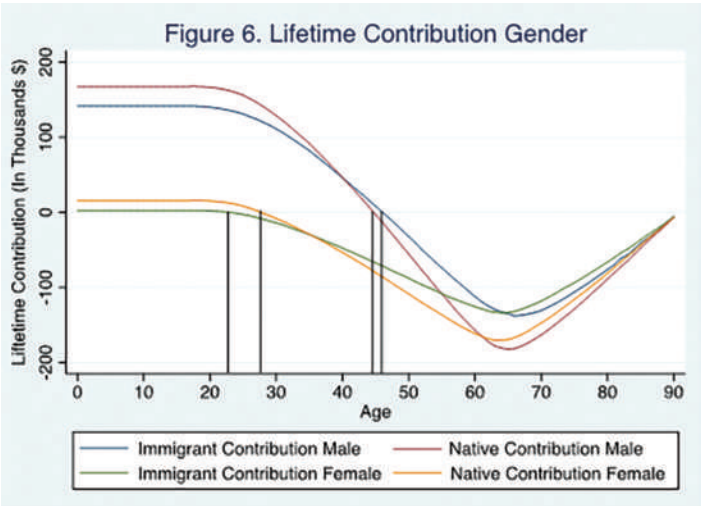
Figure 5 reveals another interesting trend that was unapparent in the yearly contribution model when comparing natives and immigrants. Although college educated immigrants seemed to reach lower peak contributions than natives during their working life, over the course of their entire life they outperform them in net positive contribution. It doesn't make intuitive sense in this context to compare the break-even point since natives don't just "arrive" in the country at a later age. However, it is still possible to draw conclusions about efficiency in impact over a lifetime by noting that it takes an immigrant arriving at age 53 or later to become a net loss while a native must start working at age 49 or before to positively contribute. Non-college educated natives and immigrants follow very similar impact trends over their lifetime, with natives' post-retirement benefits outweighing their higher earnings during their working life, leading to very similar total impacts of -$66,000 and -$75,000, respectively.

Lifecycle Model 1 - Gender

Lastly, I expand upon the trends seen in the yearly contribution gender model by looking at the impact gender has on lifetime government budget impact. As mentioned earlier, discriminating due to gender may not be the right policy decision, but understanding the life-cycle paths of both men and women can be useful in forecasting household impacts. This is especially relevant for a country with an immigration policy like the United States, which places a large emphasis on allowing family members to re-unite through immigration (Borjas 1996).

Figure 6 follows the same method as the education and average lifecycle models, this time separating yearly net contributions based on gender, and summing up each age into a lifetime impact. Figure 6 demonstrates the lifetime disparity in contributions between immigrant men and women. Assuming a working life starting at 15, men, on average, will contribute $141,000 while women, on average, will basically break even at $2,500. It is interesting to note the parabolic shape of the lifetime impact curve, which indicates that





Economics

the worse age at which an immigrant can arrive is 64 for women and 66 for men.

Figure 6 also compares the gender impact between immigrants and natives which allows us to conclude that both native men and women, on average, have a greater lifetime impact, assuming they begin working at age 15. Having seen that, on average, immigrants perform worse than natives, it makes sense that both immigrant men and women perform worse than natives. Since educated immigrants, on average, perform better than natives, however, then it must mean that one of two things (or both) could be happening. Either there are less educated immigrants than natives in my sample, pulling the immigrant averages down, or there is an omitted effect, such as the duration of stay.

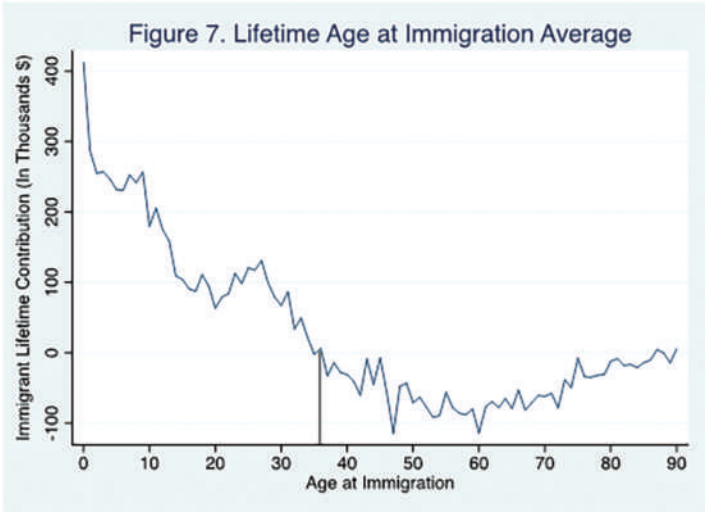
My sample data reveals that 38% of immigrants have at least one year of college education and 37.7% of natives have the same educational attainment. Therefore, the ratio of college educated to non-college educated is irrelative in this context. Earlier, we explored the age at immigration trends which revealed that, on average, immigrants have started to come in their late twenties, early thirties. I believe that the "average" immigrant's life-cycle impact is picking up this effect of later arrival in one's life.

Age at Immigration Lifecycle Model

In my second version of the life-cycle model, I test the hypothesis that earlier immigration and longer duration in the United States positively impacts earnings and therefore an immigrant's net contribution. My belief is that a 35 year old immigrant who arrived at age 1 has a different impact on the government budget than a 35 year old immigrant that arrived at age 35. The topic of many studies, it is continually shown that the earlier an immigrant arrives to the United States, the higher the likelihood that their earnings will surpass those that arrive later in their life. This phenomenon has been attributed to different factors, with higher cultural assimilation and education attainment two of the most important (Sandford Seeborg 2003)(Hoyt Chin 2010)(Schaafsma Sweetman 2011). In the same vein, the earlier an immigrant arrives, the longer he/she can positively contribute through lengthened working life taxation.

Lifecycle Model 2 - Average

The basis for my second life-cycle model is very similar to the previous one. The main difference is that I restrict the sample for each age of arrival. To do this, my first step is constructing the variable ageimmig. For each observation, I take the year at which they immigrated and subtract it from the year in which they responded to the survey. I then build a loop that runs through each age, starting at 0 and ending at 90. For each iteration of the loop, I restrict the sample to that loop number's age. Similarly to previous models, my next step is calculating the average yearly contribution at each age. This time, having a restricted sample means that I am averaging the yearly contributions for only those that immigrated at age x. I then sort my data from oldest to youngest, and sum each age to arrive at a lifetime impact. The lifetime contribution of arriving at age x is the only value that I am interested in since I am calculating that same impact for each different age of arrival. With this in mind, for each iteration of the loop, I only save that one observation. Running another loop, I append the lifetime contribution for each age of immigration to build my final model. Each observation in this model shows the lifetime impact of immigrating at age x, and unlike previous models, it accounts for the time spent in the United States. The results are plotted in Figure 7.



There are a few different results to focus on from this graph. The most important one, this time more robust than in previous models, is the age of immigration where working life contributions will offset retirement benefits received. Figure 7 suggests that any immigrant arriving after the age of 34 will, on average, have a total negative impact on the government budget if they remain in the United States throughout the remainder of their life.

The second result, which confirms my hypothesis regarding the positive effects of arriving early and having a longer duration of stay, can be observed by looking at the trends of immigrants arriving between age 0 to 15. Since all immigrants arriving in that age range start working at age 15, in theory, if there is no effect of duration of stay or earlier arrival, they should all have the same lifetime impact. Clearly, Figure 7 illustrates the opposite – a child immigrating at age 1 will contribute on average roughly $287,000 while a child immigrating at age 10 will contribute on average only $179,000. These findings are in line with (Schaafsma Sweetman 2011) and (Hoyt Chin 2010).

Thirdly, there is one interesting "spike" around age 20-27 (there are a few others around age 7 and age 45 but those qualify more as noise than structural trends). Why is it that immigrants arriving between the ages of 20 and 27 incrementally begin earning more while the overall trend shows that no matter what age you arrive, the older you are, the lower your total impact? We can't draw any conclusions about this phenomenon from Figure 7 but subsequent variations to the model will attempt to answer this question. My theory is that those arriving between 20-27 have a higher likelihood of having completed college, are looking for work abroad, and therefore are, on average, more educated than those coming in around 18 years old.

Fourth, as we saw with previous lifetime contribution models, Figure 7 also highlights a parabolic shape to the age of arrival curve. We can see that (ignoring the noise spike around age 44), the curve levels off around age 60 before rising again. This result makes intuitive sense since 60 is right around the age of retirement. If an immigrant arrives around that age and never contributes, only receiving benefits from the government, the later the immigrant arrives past that point, the smaller his negative impact will be on the government since his life expectancy will decrease.

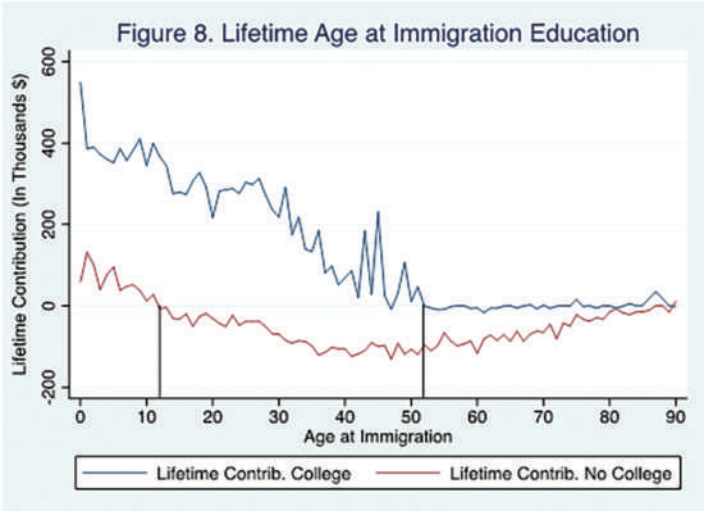
Lastly, I want to point out that we are dealing with age of arrival observations across many different years so taking averages mitigates the effects of economic downturns or booms in certain periods. It is also important to note that Figure 7 has more noise than previous models. This is in part due to the sampling size I am using and having further restricted each data point to observations for that age of immigration only.

The age at immigration trends discussed earlier revealed that immigrants have, over the past five years, been coming in to the United States around age 28. From an immigration policy standpoint, it is interesting to see the lifetime impact that the average 28-year-old immigrant has on the government budget amounts to roughly $99,500. Although the graph has not been included, it is also interesting to see that a 27-year-old man has a $234,000 lifetime impact and a 29-year-old woman has a $25,600 lifetime impact, on average.

Lifecycle Model 2 - Education

I continue to build on variations to my second life-cycle model by evaluating the impact trends due to differing educational attainment. Comparable to my hypothesis regarding the effects of earlier arrival, I also believe that the effect of education will shed more light on the effect of age at arrival. Simultaneously, age at arrival has an impact on educational attainment, as a younger immigrant will have an easier time assimilating to the culture, therefore giving him/her a better chance at higher human capital accumulation (Schaafsma Sweetman 2011)(Myers et al. 2009). Once again, I run the same loops as discussed in the previous section, this time separating my sample into those with at least one year of college education and those with less.

In Figure 8 we can see that earlier arrival evidently has an impact on lifetime contribution. Restricting once again our analysis to those arriving between ages 0 and 15, however, we don't observe the same major differences in lifetime contributions for those college educated. While age 0 seems to be an outlier, ages 1 to 14 all oscillate in the earnings ranges of $345,000 to $400,000. What could be the reason for not seeing a clear impact of earlier arrival? This is most likely because they all end up with a college education. The impact of earlier arrival may be increased educational attainment, manifested in the oscillation observed in Figure 8. The homogeneity in my sample would simultaneously suggest that earlier arrival, age 1 versus age 10, affects future lifetime contribution by affecting the chances of achieving a college education. Therefore, when looking at a sample of only college educated immigrants, the effect of earlier arrival (before working age) is non-existent.



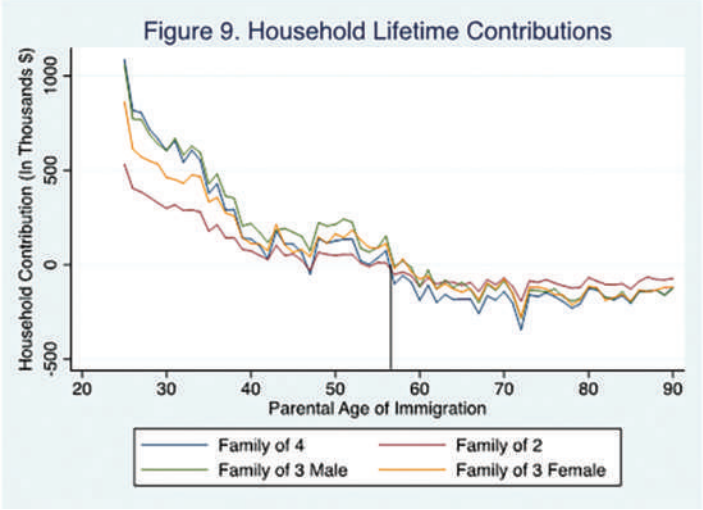
Around age range 40-47, we observe a significant spike most likely due to highly-educated immigrants, those with PhDs, arriving at those ages. Another interesting feature of Figure 8 is the lifetime impact cutoff point of $0, which takes place at age 52; this is much older than the 34 years of age suggested for the "average" immigrant. This lifetime graph is also very particular in that it doesn't have the same parabolic shape as all the previous ones. Rather, past age 52 it oscillates around $0. I do not have any intuitive theory as to why this takes place, however, my best guess is that older college educated immigrants coming to the United States are most likely wealthier. Instead of coming for work or medical care, they come to settle and retire – creating minimal impact on the government budget. Additionally, I addressed this data concern in the previous section, and it may have an even stronger impact since my sample has greater restrictions, but my results may be in part due to a lack of enough observational data. This may also explain the increased noise in the college-educated data relative to the non-college educated.

For non-college educated immigrants, the earlier arrival effect does take place. Within that sample of immigrants, arriving at age 1 makes a significant difference from arriving at age 10. Since their educational attainment remains low throughout their life, a potential theory might be the effect that the language barrier has on future earnings – the later they arrive, the stronger the effect. The most puzzling aspect of Figure 8 for non-college educated immigrants is that the lifetime impact $0 cutoff takes place at age 12. Arriving at age 12 versus arriving at age 13, unlike arriving at age 24 instead of age 25, will not increase your net contribution by an extra year since you may only start working at age 15. This provides even stronger evidence that there must be some significance in arriving at an earlier age, even prior to working age eligibility. Unfortunately, I am unable to explain why the parabolic shaped curve's inflection point takes place around age 45. The only theory I could have for a 55-year-old making less of a negative impact on the government budget than a 45-year-old is if they both come in, receive benefits right away, and don't find work. The plausibility of a working age foreigner receiving benefits upon immigration is relatively low, however.

Household Lifecycle Model

Immigrants who come to the United States are often not alone - they come with families. I test my hypothesis that the impact of a family on the government budget differs from the impact of a sole immigrant through the construction of a household lifecycle model. This model helps me better approximate the impact of children on a household while varying the different types of household possibilities.

I construct four different types of immigrant households: 1) a parent and child, 2) a father, son, and daughter, 3) a mother, son, and daughter, and 4) a father, mother, son, and daughter. I assume that the parents have children at age 25 and that once the parents die the children keep contributing until their own death. For each household, I attempt to answer the question, "What is the contribution of a household when the parent(s) immigrate(s) at age x



with child(ren) aged x-25?". More importantly, the answer to this question helps shed light on the optimal immigration policy when considering multi-person households. I plot my results in Figure 9.

To project the 2-person household's net contribution, I sum the contribution of a parent immigrant arriving at age x, with their child's contribution being an immigrant arriving at age x–25. Doing this for every x>24 provides a full view of all age possibilities at parent arrival. I construct the 3-person households by summing the contribution of either a male or female parent arriving at age x>24 with their son and daughter arriving at age x-25. Similarly, to construct a 4-person household I sum the contribution of a male immigrant and female immigrant arriving at age x>24 with a male immigrant and female immigrant arriving at age x-25.

My results show that a 2-person household will positively impact the government budget over their lifetime if the parent arrives before age 57. Similarly, a 3 and 4-person household will also positively impact the government budget as long as the parents arrive prior to age 57. The shape of the household lifecycle curves reveal that a 3-person male parent household, on average, will have a greater positive contribution on the government budget if the parents arrive prior to age 57, however, the 4-person household will have a greater negative contribution if they arrive past that age.

The 4-person household follows a very similar pattern to the 3-person male parent household, revealing that a female spouse does not seem to have much of an impact on the government budget. In fact, around parent arrival age of 30-57 the 3-person male parent household shows a stronger positive impact than the 4-person household. The 3-person female parent household and 2-person household both have less positive and negative overall impacts on the government budget. A 3-person household led by a male shows a stronger positive impact prior to arriving at age 57 than one led by a female, however, past age 57 the magnitudes of their impact are very similar. It is also important to note that around age 47, the 3-person households both begin to have a more positive impact than the 4-person household. This is likely due to the parents hitting retirement age sooner and with 2 parents rather than 1 in the household, the negative impact is amplified.

Interestingly, all curves follow a slight parabolic shape whereby a parent arriving around age 72 will have the worse impact on the government budget than if they came at an earlier or later age. This may be explained by the amount of time that both the parent and child will spend receiving aid from the government past retirement age. A parent coming in at age 72 can expect to receive welfare until death, while their child at age 47 will contribute less over their working life than when they hit retirement. Past that point, the older households will not receive welfare for as many years and therefore will have less of a negative impact.

Conclusion

Limitations

One of the main limitations of my paper is that it ignores the impact of immigrants on native wages and employment displacement effects, and instead relies purely on estimating their fiscal impact. The argument is usually made that immigrants are displacing native workers when they gain employment, which would mean that the taxes paid for the job displaced does not change whether an immigrant or native holds the position. Under my model it would seem as if the immigrant is benefitting the government by the total amount of taxes paid. Additionally, while the immigrant may not receive welfare, the now displaced worker may begin receiving welfare from the government, creating a net loss.

Another potential limitation in my paper is the lack of a death discount factor, which would better predict lifetime impacts by discounting it based on life expectancy. According to the Social Security Administration's Actuarial Life Table, male life expectancy ranges roughly between 75-80 years old while female life expectancy ranges between 80-85. These ranges are very similar to where my data cuts off, which mitigates the effects of a large impact on my results.

Lastly, this paper does not account for the public cost of education. My results show that the younger an immigrant arrives, the higher the lifetime contribution. However, while an immigrant arriving prior to schooling age is shown to have a greater positive fiscal impact than one that may arrive at age 25, for example, the cost of education could impact the net impact. If both immigrants end up having the same level of schooling, one's native country will have paid for it while the other will have it paid for by the United States. It makes it extremely difficult to compare such a scenario since other factors such as cultural integration can have a large impact on future success, regardless of whether two immigrants have the same level of education.

Policy Implications

To better understand policy implications of the results displayed in this paper, it's important to understand the current U.S. Immigration Policy Goals. These revolve around 4 axes: 1) Economic: increase labor supply, especially where skill deficits exist, 2) Humanitarian: reunite families, 3) Cultural: ethnic and racial diversity, 4) Political: allowing or refusing certain political refugees. This contrasts to the Canadian immigration "points system", which allocates visas to prospective immigrants by awarding points based on language proficiency (28), education (25), experience (15), age (12), employment contract (10), and adaptability (10).

My results clearly indicate that the United States should actively solicit young, highly-skilled immigrants. Similar to Canada, my results for the average immigrant point to the optimal working ages of 18-35, my cutoff point showing age 35 as the beginning of

a net loss. The only difference being that they award no points for immigrants below the age of 18, while my results all indicate that the younger the immigrant arrives, the higher the lifetime contribution.

Based on the two main metrics observed, age at arrival and education, my results suggest that immigration can be a solution to the aging population and fiscal deficit. If the U.S. goal of immigration aims to solve this growing problem, the government has a clear incentive to implement a similar "points" system to Canada. Other potential government interventions could include ways to reform social benefits provided to immigrants through optimal time-dependent structures. These would ensure that they are making a net positive impact on the system over their lifetime, while also considering their impact on the native population.

Final Remarks

This paper demonstrates that immigrants have strong quantitative implications for fiscal policy in the United States. In particular, this paper investigates the optimal lifetime contributions based on age at arrival, education, and gender. The findings throughout the paper are illustrated by computing the net fiscal impact, in present value terms, of admitting one additional immigrant to the United States, conditional on education, gender, and age at time of arrival. The lifetime contributions vary considerably across these three characteristics, with large and positive values for college-educated immigrants arriving in the earlier part of their life.

Using a yearly net contribution model, two life-cycle models, and a household contribution model, I demonstrate that the average immigrant arriving past age 34 has a lifetime negative fiscal impact. Additionally, a college educated immigrant arriving prior to age 52 will have a lifetime positive fiscal impact while a non-college educated immigrant will roughly have a lifetime negative fiscal impact, regardless of age at arrival. Further, I confirm that age at arrival matters, and determine that arrival prior to working age influences educational attainment. Finally, I provide a household life-cycle model that sheds light on the fiscal contribution of immigrating families.

My research has a few avenues for add-on research. First, I do not distinguish between the level of welfare and taxes for each individual's contribution. Understanding that dynamic might better motivate welfare or tax reform. Second, I focus on the mix of immigrant characteristics, but not on the level of immigration. Third, while I have data on education, age, and gender, a further analysis could look at the effects of language proficiency, skill, and native country of birth (Lagakos 2016). Lastly, it is important to note that while these results are based on U.S. immigration policy, there is an external validity concern, as each country has its own unique welfare system policies that also differ in magnitude.

References

Auerbach, Alan J., and Philip Oreopoulos. "The Fiscal Effect Of U.S. Immigration: A Generational-Accounting Perspective". Tax Policy and the Economy 14 (2000): 123-156.

Bleakley, Hoyt, and Aimee Chin. "Age At Arrival, English Proficiency, And Social Assimilation Among US Immigrants". American Economic Journal: Applied Economics 2.1 (2010): 165-192.

Borjas, G. J., and L. Hilton. "Immigration And The Welfare State: Immigrant Participation In Means-Tested Entitlement Programs". The Quarterly Journal of Economics 111.2 (1996): 575-604.

Borjas, George. "Immigration And The American Worker: A Review Of The Academic Literature". Center for Immigration Studies (2013).

Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren. Integrated Public Use Microdata Series, Current Population Survey: Version 4.0. [dataset]. Minneapolis: University of Minnesota, 2015. http://doi.org/10.18128/D030.V4.0.

Friedberg, Rachel M, and Jennifer Hunt. "The Impact Of Immigrants On Host Country Wages, Employment And Growth". Journal of Economic Perspectives 9.2 (1995): 23-44. Web.

Gibson, Campbell, and Emily Lennon. "Historical Census Statistics On The Foreign-Born Population Of The United States: 1850 To 1990". Working Paper no.29, U.S. Census Bureau (1999).

Hansen, Jorgen, and Magnus Lofstrom. "Immigrant Assimilation And Welfare Participation: Do Immigrants Assimilate Into Or Out Of Welfare?". The Journal of Human Resources 38.1 (2003): 74.

Klopfenstein, Kristin. "A Reconsideration Of Immigrant Assimilation And Welfare Participation". (1998).

Lagakos, David. "Life-Cycle Human Capital Accumulation Across Countries: Lessons From U.S Immigrants". National Bureau of Economic Research 21914 (2016): n. pag. Print.

Myers, Dowell, Xin Gao, and Amon Emeka. "The Gradient Of Immigrant Age-At-Arrival Effects On Socioeconomic Outcomes In The U.S". The International Migration Review 43 (2009): 204-229.

Ottaviano, Gianmarco I. P., and Giovanni Peri. "Rethinking The Effect Of Immigration On Wages". Journal of the European Economic Association 10.1 (2011): 152-197. Web.

Pekkala Kerr, Sari, and William R. Kerr. "Economic Impacts Of Immigration: A Survey". SSRN Electronic Journal (2011): n. pag. Web.

Peri, Giovanni. "Highly-Educated Immigrants And Native Occupational Choice". (2010).

Sandford, Jeremy, and Michael C. Seeborg. "The Effects Of Ethnic Capital And Age Of Arrival On The Standard Of Living Of Young Immigrants". Journal of Economics 29 (2003): 27-48.

Schaafsma, Joseph, and Arthur Sweetman. "Immigrant Earnings: Age At Immigration Matters". Canadian journal of Economics (2011): 1066-1099. Print.

Storesletten, Kjetil. "Sustaining Fiscal Policy Through Immigration". Journal of Political Economy 108.2 (2000): 300-323. Web.

"The Fiscal Impact Of Immigration In OECD Countries". International Migration Outlook (2013).

Zhang, Yi. "A Reconsideration Of Labor Supply Of Immigrants And Social Welfare Programs". State Univeristy of New York at Stony Brook (2014).

A Review of the Biological Aspects of Autism Spectrum Disorder

Jennifer Bi '18, Puneet Gupta '17, Katherine Miclau '20, Sarah Ryan '20,
Yong Shen '19, Katherine Venturo-Conerly '20, Akash Wasil '19

Autism Spectrum Disorder (ASD) has increasingly become an important topic of research due to its status as a major public health concern. It exerts immense suffering on children and families and imposes an economic burden on society. As suggested by its name, ASD does not manifest in one form or present only a single set of symptoms, but instead contains heterogeneous sets of symptoms, leading to obstacles that obstruct the path to ASD etiology discovery. Under the DSM-5, a diagnosis of ASD requires deficits in social skills, communication, and repetitive behaviors (McPartland, Reichow, & Volkmar, 2012). Traditionally, researchers have focused on identifying genes that underlie most cases of ASD and have discovered several genes that correspond to a wide range of ASD cases. This approach, however, has limitations as ASD is now understood to have a much more complex set of causes than the known set of genes. In light of evidences suggesting the complexity of ASD etiology, some researchers have shifted focus from finding a universal cause for all symptoms of ASD to systematically grouping symptoms and comorbidities, and performing genetics studies to discover a set of genetic factors contributing to these groups of symptoms. Other researchers tackle the genotypic heterogeneity by attempting to identify genetic factors underlying phenotypically homogenous subgroups of ASD. Moreover, some researchers have broadened the scope of ASD etiology investigation to include the investigation of comorbidity etiology. This review elucidates the current strategies in discovering the biological basis of ASD research and highlights some of the most recent advancements in the understanding of ASD etiology.

Awareness and research relating to mental disorders has increased drastically in the past couple decades. Today, there are a multitude of mental health disorders that have gained importance in the clinic. One of the more prominent, though not completely understood, mental health disorders is autism spectrum disorder (ASD). The term autism spectrum disorders (ASD), also known as pervasive developmental disorders (PDD), was developed in 2013 when the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders was released. ASD functions as an umbrella term for developmental disorders involving primarily communication, behavioral, and intellectual issues, such as autistic disorder and Asperger syndrome.[1] The Center for Disease Control and Prevention (CDC) estimated in 2016 that the prevalence of ASD was 1 in 68 children aged eight years, and more specifically, 1 in 42 for boys and 1 in 189 for girls.[2] These numbers have significantly increased ever since initial studies on ASD prevalence were done in the 1990s due, in part, to increased awareness and improved screening. Currently, ASD is often difficult to diagnose in clinics due to heterogeneity in symptom profiles within ASD, comorbidities between ASD and other diagnoses, and symptom overlap between ASD and intellectual disability. Early signs may start to appear as early as infancy, but many parents are unsure of whether these signs reflect ASD or non-pathological personality development.[3] However, as a child ages, these early signs may develop into noticeable differences in social behavior and communication.[4]

ASD has been linked to many families and groups of genes, and new discoveries are continually being made. Though the attempt to find a single underlying gene for ASD has not been determined, research focused on specific DNA segments and proteins has been initiated. Today, researchers have begun subgrouping ASD based on genotypic uniformity and are trying to understand the etiological heterogeneity on ASD and its comorbidity.

Comorbidity and Etiology

ASD is commonly comorbid with other neurodevelopmental or behavioral disorders. Thus, a focus on one of these disorders may cause a physician to neglect the consideration or diagnosis of another. More specifically, it may result in a failure to diagnose autism.

In particular, efforts to subdivide autism and do away with the confounding phenotypic heterogeneity observed in the disorder have become increasingly focused on the presence of ASD comorbidities. Earlier research has shown that ASD has a large variety of associated comorbidities, including gastrointestinal disorders, sleep disorders, epilepsy, psychiatric illness, and immune disorders. Recently, an attempt to stratify ASD using comorbidity clusters has been made, leading to the idea that ASD with one group of comorbidities should be viewed somewhat differently from ASD with another group of comorbidities.[5]

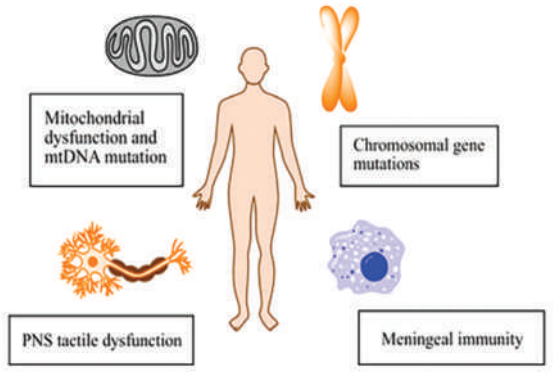


Figure 1. Comorbidities and possible factors contributing to the development of ASD.

In order to find the optimal groups, researchers analyzed the clinical trajectories of a large group of ASD patients. They examined clusters of forty-five common comorbidities in around 5,000 ASD patients, checking in with the ASD patients every six months from birth to age fifteen. The clusters they identified showed distinct medical trajectories, i.e. they followed similar developmental paths towards similar phenotypes. The three focal clusters identified were mainly characterized by seizure, psychiatric disorders, and gastrointestinal disorders, along with other minor comorbidities. The implication is that there might be distinct etiological roots that lead to ASD with seizures, roots that lead to ASD with psychiatric illnesses, and roots that lead to ASD with gastrointestinal disorders. Therefore, these clusters can be seen as subgroups of autism with distinct comorbidities and possibly distinct causes. This approach to subgrouping autism does not rely on previously identified factors, yet it still manages to create homogenous subgroups. Certainly, this approach has great potential in contributing to the search for causative factors of ASD.

Moreover, ASD is also commonly associated with delays in language development. However, many of the language development issues that people with ASD experience are not specific to ASD, and also often occur in people with developmental disabilities and/or other types of disorders. Genetic studies show overlap between specific areas of the chromosome that seem to be associated with both the expression of Specific Language Impairment (SLI) and the communication problems that arise in ASD patients. ADHD is another example of a disorder that is commonly comorbid with ASD. One study in particular showed that the ASD + ADHD group had lower marks on the Vineland Adaptive Behavioral Scales (VABS-II) and the Pediatric Quality of Life Inventory (PedsQL). These results suggest that the presence of both ADHD and ASD contributes to a lesser quality of life due to a hindered ability to execute skills needed in everyday life.[7] Moreover, in areas of social interaction, the presence of both ADHD and ASD results in language impairment, executive functioning difficulties, and learning disabilities. Thus, children who are on the spectrum are also commonly suggested to screen their children for both ASD and ADHD.

Research suggests that the co-occurrence may be the result of two separate disorders with a common etiology. That is, the two disorders may share a common genetic basis.[7] To further support this theory, both twin and family studies were undertaken, and the results showed that ASD is highly genetic. Linkage studies and Genome Wide Association Studies (GWAS) have pointed to pleiotropic genes, loci, and single nucleotide polymorphisms (SNPs). Pleiotropic genes are defined as one gene affecting multiple phenotypic traits, and SNPs are defined as a variation in a single nucleotide occurring at a specific position on the genome.[8] Thus, the gene has more than one allele that can cause variations in the amino acid sequence. Various brain imaging scans have also been conducted, but specialists agree that solely looking at the brain is a simplification of the problem.

Studies and experts alike have shown the co-occurrence of ASD and other disorders result in more symptoms and a more complex prognosis. However, oftentimes only one diagnosis is given. As a result, ASD can go undiagnosed or untreated. Therefore, a better genetic and biological understanding of how ASD and other disorders interact is needed in order to give patients a proper diagnosis of ASD.

Etiological heterogeneity in ASD

Many research studies have explored the roles of rare mutations and genetic imbalances in Autism Spectrum Disorder (ASD). A series of various studies of cytogenetics and whole-genome linkage of exome sequencing has shown that the etiological component of autism is extremely complex and interwoven, presenting a high degree of pleiotropy and locus heterogeneity.[9]

ASD has a strong genetic basis revealed by the recurrence risk in families, twin studies, and the co-occurrence with chromosomal disorders and rare genetic syndromes. The information involving biological origins autism and mechanisms continues to be updated and become more extensive with the progression of genetics and more conducted animal model systems. Currently, ASDs are diagnosed in about 1% of children, four times more common in males than in females.[10]

Though scientific evidence is relatively scarce, some evolutionary psychologists have identified traits in autistic individuals that may have been beneficial. These traits, such as restricted interests, could lead to innovation and have been linked to the maintenance of autism alleles in the gene pool. Recent studies have speculated that as many as 1000 genes are implicated, including both rare mutations (occurring in less than 5% of the general population) with significant effect sizes as well as common variants (occurring in more than 5%) with smaller effect sizes.[9] Rare mutations with minor allele frequency are often identified in the form of chromosomal abnormalities (present in 5% of autistic individuals), Mendelian genetic syndromes (i.e. syndromic autism in 5%), rare copy number (5-10%), and de novo mutations (5-10%). Common variants, on the other hand, could also contribute to the emergence of autism: it has been estimated that as many as 40% of simplex families and 60% of multiplex families could have increased autistic traits due to single nucleotide polymorphisms.[9] De novo mutations, which are microdeletions, microduplication and single nucleotide variants, occur in particularly high rates in the paternal germline copy number mutations. Many concurrent genomic mutations have been proved to lead to expression of autism, however each of these individual factors alone is only minutely responsible for ASD in and of itself. For example, each copy number variation is only present in at most 1% of individuals with autism, revealing the very significant role of genetic heterogeneity.

Both past and current research has shown that there are a multitude of genetic implications involved in ASD. Due to the complex nature of the etiology of autism, a total of 103 disease genes and 44 genomic loci have been linked with ASD or autistic behavior. Only 10%–20% of autistic individuals have an identifiable genetic etiology, with 5% of cases presenting visible chromosomal alterations, the most frequent of which include 15q11–q13 duplications, and 2q37, 22q11.2 and 22q13.3 deletions. ASD can also be due to mutations of single genes involved in X-linked disorders, autosomal dominant and recessive, the most frequent being fragile X syndrome in 2% of cases.[11] Mutation in the mitochondrial DNA (mtDNA) constitutes another portion of the possible contributing factors of ASD. Recent research has shown the positive correlation between mitochondrial dysfunction and autism.[11] The incidence of mitochondrial disease occurring in ASD population is about 5%, whereas in the general population that number is roughly 0.01%.[12] A recent analysis of mtDNA sequences from ASD children and their unaffected siblings found a pattern of heteroplasmic mtDNA mutations related

to increased risks of developing ASD.[13]

Finally, the etiological heterogeneity of ASD is further complicated by the plethora of associated comorbidities and recent evidence that suggests some comorbidity might actually be important factors in ASD development. Recent research showed that interferon-γ, a chemokine secreted by immune cells, is necessary for proper mouse social behavior.[14] A single injection of interferon-γ was able to restore social preferences in the mouse, and decrease hyperconnectivity in the prefrontal cortex, one of the most common characteristics of ASD.[14] Disorders in the immune system, one of the comorbidities of ASD, could actually influence the development of ASD. This is not completely surprising as the immune system has been shown to influence memory formation and learning as well.[15] For instance, inflammation has been shown to lead to changes in microglia. These changes can overprune synapses in patients with autism.[16] Furthermore, peripheral nervous system (PNS) disorder is found to play a role in the development of ASD. A common comorbidity of ASD is abnormal tactile sensitivity. Recent research shows that the mutation of genes *Mecp2*, *Gabrb3*, *Shank3*, and *Fmr1* in mouse cause altered tactile sensitivity.[17] Deletion of *Mecp2* or *Gabrb3* during early development specifically in somatosensory neurons causes social interaction deficits and anxiety-like behavior, while disruption of the same genes during adulthood leads to less severe ASD phenotypes. Rescue of *Mecp2* null mutant mice specifically in somatosensory neurons with functional *Mecp2* during early development restores social interaction deficits and reduces anxiety-like behaviors.[17] The research suggests that tactile dysfunction, a comorbidity of ASD, could directly contribute to ASD development.

One Gene Utopia

There is currently no known biomarker or specific genomic sequence that can be linked to ASD as the main causative factor. However, in a genetic utopia, clinical and laboratory researchers would be able to find a genomic sequence or biomarker common to all ASD patients. For this purpose, researchers have recently explored four particular genes and proteins that are deemed as high risk factors for ASD, although these mutations are not present in all ASD patients.

One gene that has been investigated recently is the *TSHZ3* gene that encodes a zinc finger transcription factor, which consists of a zinc finger-binding domain that allows it to bind DNA, RNA, and other proteins. The *TSHZ3* gene has been shown to be one of the genes most highly expressed in the developing human neocortex, however, its exact function is not well understood. In recent research, it was found that mice heterozygous for the *TSHZ3* gene were affected and displayed a change in function of the synapses that are found between cerebral cortical projection neurons.[17] Not only was synapse function affected, but also the mice displayed abnormalities in behavior similar to ASD patients.[18]

Chromodomain helicase DNA binding protein 8 (*CHD8*), which encodes a chromatin remodeling protein, is another gene that is commonly affected by mutations in patients with ASD. The protein encoded has been shown to be involved in regulating the p53 pathway and *CTNNB1* gene, and it interacts with CHD7 protein, which is involved in many human abnormalities and defects.[19] Recently, a study showed that mice that were heterozygous for *CHD8* mutations demonstrated characteristics similar to those of ASD patients, such as increased anxiety, altered social behavior, and

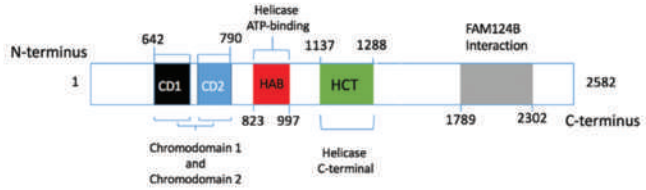


Figure 2. CHD8 protein structure in humans. Diagram of the structure of the CHD8 protein with indications of different domains and motifs. Numbers positioned above and below indicate the amino acid positions.

repetitive behavior.[20] Furthermore, they showed that neurodevelopment was delayed in the mice with haploinsufficiency mutation of *CHD8*. Moreover, this haploinsufficiency was found to be associated with the activation of Re-1 silencing transcription factor (REST), in both mice and humans. Activation of REST leads to suppression of the transcription of many neuronal genes, thereby delaying development.

Another protein that was recently studied is the histone acetyltransferase CREB binding protein. Mutations in the protein were found to be associated with a large group of ASD human patients after a genomic study. In particular, this protein is involved in chromatin regulation and DNA function, and controls gene expression by its role as an epigenetic regulator that modifies chromatin. In mice, it was found that those with a deletion mutation in the CH1 domain of the protein displayed behaviors similar to that of ASD patients, including hyperactivity, social interaction deficits, motor dysfunction, abnormal synaptic plasticity, impaired recognition memory, and repetitive behavior, indicating that this CBP protein is likely a high risk factor in autism.[21]

The fourth gene of interest is the *GABRA5* gene, which encodes α5 $GABA_A$ (gamma-aminobutyric acid type A) receptors. Upon deletion of this gene, autism-associated behaviors were exhibited in mice. Under limited social contact in certain tasks, mice also exhibited memory deficits. Moreover, they also showed excess repetitive behaviors and impaired cognitive function/problem solving abilities, which were measured through self-grooming and their reactions to a puzzle box, respectively. In analysis of 396 human cases of ASD, two rare missense variants were found in the *GABRA5* gene.[22] These receptors were also found to be previously downregulated and show low mRNA and protein levels in ASD patients.[23,24]

Given our understanding of the research advances made in these four different genes and proteins indicated as risk factors in ASD, it is clear that the underlying cause of ASD cannot yet be implicated in a specific biomarker or gene. Many of the genes and proteins that have been found to be involved in ASD cases are usually found to be involved in other issues and diseases as well, making ASD a difficult disorder to categorize and treat.

Subgrouping ASD based on genotypic uniformity

One approach that researchers have taken to tackle this apparent genotypic heterogeneity is subgrouping ASD in such a way that each genetic problem corresponds to a phenotypically homogenous subgroup of ASD. These studies generally attempt to correlate specific genotypes to homogenous phenotypic expressions. One study assumed phenotypic heterogeneity arises from genotypic hetereogeneity of ASD and homogenous genotype causes phenotypic

homogeneity.[25] In accordance with this hypothesis, the group isolated a group of ASD patients with a specific genetic mutation implicated in ASD, 22q11 deletion syndrome, and compared the phenotypic homogeneity within that group to the phenotypic homogeneity in a mixed group of ASD patients. They found that the group with the same genotype had a much greater phenotypic homogeneity than that of the mixed group. These results imply that clinical heterogeneity of ASD can be reduced by subgrouping ASD patients based on specific genotypes.

However, as studies discover more genes implicated in ASD behaviors, it is important to recognize the limitations of such findings. In another study, the authors highlight how ASD is not a single clinical entity but instead a behavioral manifestation of anywhere from tens to hundreds of genetic and genomic disorders.[9] Thus, the behavioral presentation of the individual should be taken into consideration when focusing on genetic mutations due to the complexity of ASD as an individually-specific disorder with various degrees of intellectual disability (ID) severity. Furthermore, autism is often associated with other neuropsychiatric disorders such as attention deficit-hyperactivity disorder, bipolar disorder, and obsessive compulsive disorder. Indeed, many of these studies indicating genetic mutations leading to ASD were conducted without indication of how formal diagnostic evaluation was performed.

The specific studies on the similar genetic causes of ID and ASD (70% comorbidity), and of ASD and epilepsy (25% comorbidity) indicate these neurodevelopmental disorders share common genetic bases. These studies redefine autism as the final common pathway for often co-occurring genetic brain disorders: well-recognized ID-genes, which do not always result in symptomatic ID, as well as some gene mutations characteristic of epilepsy can also correlate to ASD.[9] Instead of considering ASD as one distinct disease, these genetic findings portray ASD as a continuum of neurodevelopmental disorders manifesting in diverse manners based on other genetic, environmental or stochastic factors. In fact, many disorders are well known to have a very high comorbidity with ASD: the most frequent are *SHANK3* mutations, tuberous sclerosis (*TSC1*, *TSC2*), Rett syndrome (*MECP2*), fragile X syndrome (*FMR1*), Smith–Lemli–Opitz syndrome (*DHCR7*), cortical dysplasia-focal epilepsy syndrome (*CNTNAP2*), and adenylosuccinate lyase deficiency (*ADSL*).[9]

Discussion and Conclusions

In order to better understand the underlying biological basis of ASD and to continue the advancement of treatment of future ASD individuals, it is essential to further emphasize the role of ASD in conjunction with other mental, neurodevelopmental or behavioral disorders. Because autism is caused by such a large variety of genetic abnormalities that cannot be linked back to a singular mutation or a uniform change in the genomic sequence consistent for all ASD individuals, it is very important to consider the nonhomogeneous nature of ASD and embrace its genomic complexities. As comorbidity is so common and widely varied for every affected individual, identifying the common genotypic base of ASD in relation to other disorders has been shown to strongly correlate to the prominent phenotypic groups in autistic patients. Focusing on subgrouping autism into more specific, phenotypically homogenous groups of ASD individuals is a promising approach to treating individuals with common symptoms and predicting future trajectories; particularly, the three most indicative clusters of recurrent comorbidities can be characterized by seizure, psychiatric disorders and gastrointestinal disorders. Continuing to identify various comorbidities into more individualized and specific subgroups will aid in furthering our understanding of the complexities of autism and exactly what characterizes it.

Research into Autism Spectrum Disorder (ASD) presents a discourse of questions and analysis concerning the biological basis and optimal treatment methods for the disorder. Although this type of research is necessary, it is also imperative to discuss and consider the sociocultural factors that play a role in ASD. These factors have significant implications for both diagnosis and treatment. When striving to understand the social and cultural basis of ASD, the hope is not to bring attention away from studying the underlying biology; rather, this research seeks to provide further insight into understanding the biology and provide a bridge between the multifaceted components and complexities of this disease.

An overwhelming amount of research makes clear one simple fact: diagnosis of autism requires *recognition* of autism. While this seems intuitive, a failed or delayed diagnosis is often a result of one of three circumstances: an inability to construct a universal, concrete framework for signs of autism, the failure to acknowledge cultural differences in defining the disease, and the tendency to attribute signs of autism to external or coexisting conditions. Any combination of these factors influences when and how someone is diagnosed with autism; whether it be by a family member, community member, or physician. Ultimately, there needs to be more emphasis on and responsibility for recognizing the disease in its various yet different forms. Autism is not a disease that can be generalized. Therefore, it is best to assess and diagnose each person individually.

References

1. Rybakowski F., Bialek, A., … and Dunajska, A. (2014) Autism spectrum disorders—epidemiology, symptoms, comorbidity and diagnosis, *Psychiatr Pol, 48*(4), 653–65.
2. Christensen D., Baio J., Braun K., … and Yeargin-Allsopp, M. (2016) Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, *MMWR Surveill Summ*, *65*(No. SS-3)(No. SS-3), 1–23.
3. Tarbox, J., Matson, Johnny L., Sturmey, P., and Dixon, D. (2014) Handbook of Early Intervention for Autism Spectrum Disorders : Research, Policy, and Practice (Autism and Child Psychopathology Series), New York, NY, Springer New York, Imprint: Springer, pg 1.
4. Lord, C., Cook, E., Leventhal, B., Amaral, D. (2000) Autism spectrum disorders, *Neuron* 28, (2), 355-363.
5. Doshi-Velez, F., Ge, Y., and Kohane, I. (2014) Comorbidity clusters in autism spectrum disorders: An electronic health record time-series analysis, *Pediatrics*, *133*(1), E54-63.
6. Leyfer, O., Tager-Flusberg, H., Dowd, M., Tomblin, J., and Folstein, S. (2008) Overlap Between Autism and Specific Language Impairment: Comparison of Autism Diagnostic Interview and Autism Diagnostic Observation Schedule Scores, *Autism Research 1*, 284-296.
7. Leitner, Y. (2014) The co-occurrence of autism and attention deficit hyperactivity disorder in children - what do we know?, *Frontiers In Human Neuroscience*, *8.*
8. Solovieff, N., Cotsapas, C., Lee, P., Purcell, S., and Smoller, J. (2013) Pleiotropy in complex traits: Challenges and strategies, *Nature Reviews Genetics, 14*(7), 483.
9. Betancur, C. (2011) Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting, *Brain Research, 1380*, 42-77.
10. Lai, W., Goh, T., Oei, T., and Sung, M. (2015) Coping and Well-Being in Parents of Children with Autism Spectrum Disorders (ASD), *Journal of Autism and Developmental Disorders, 45*(8), 2582-2593.
11. Giulivi, C., Zhang, Y., Omanska-Klusek, A., and Pessah, P. (2010) Mitochondrial Dysfunction in Autism Cecilia, *Jama*, *304*, 2389–2396.
12. Rossignol, D., and Frye, R. (2012) Mitochondrial dysfunction in autism spectrum disorders: a systematic review and meta-analysis, *Molecular Psychiatry*, *17*(3), 290–314.
13. Wang, Y., Picard, M., and Gu, Z. (2016) Genetic Evidence for Elevated Pathogenicity of Mitochondrial DNA Heteroplasmy in Autism Spectrum Disorder, *PLOS Genetics*, *12*(10), e1006391.
14. Filiano, A., Xu, Y., Tustison, N., Marsh, R. Baker, W., Smirnov, I., … and Kipnis, J. (2016) Unexpected role of interferon-γ in regulating neuronal connectivity and social behaviour, *Nature*, 1–18.
15. Derecki, N., Cardani, A., Yang, C., Quinnies, K., Crihfield, A., Lynch, K., & Kipnis, J. (2010) Regulation of learning and memory by meningeal immunity: a key role for IL-4, *J Exp Med*, *207*(5), 1067–1080.
16. Orefice, L., Zimmerman, A., Chirila, A., Sleboda, S., Head, J., & Ginty, D. (2016) Peripheral Mechanosensory Neuron Dysfunction Underlies Tactile and Behavioral Deficits in Mouse Models of ASDs, *Cell*, *166*(2), 299–314.
17. Pardo, C. A., Vargas, D. L., & Zimmerman, A. W. (2005). Immunity, neuroglia and neuroinflammation in autism. *International Review of Psychiatry, 2005, Vol.17(6), p.485-495, 17*(6), 485–495.
18. Caubit, X., Gubellini, P., Andrieux, J., Roubertoux, P. L., Metwaly, M., Jacq, B., … and Fasano, L. (2016) *TSHZ3* deletion causes an autism syndrome and defects in cortical projection neurons, *Nature Genetics*, *48*(11), 1359–1369.
19. Krumm, N., O'Roak, B., Shendure, J., and Eichler, E. E. (2014) A *de novo* convergence of autism genetics and molecular neuroscience *Trends in Neurosciences*, *37*(2), 95–105.
20. Katayama, Y., Nishiyama, M., Shoji, H., Ohkawa, Y., Kawamura, A., Sato, T., … and Nakayama, K. (2016) CHD8 haploinsufficiency results in autistic-like phenotypes in mice, *Nature,* 537(7622), 675-679.
21. Zheng, F., Kasper, L., Bedford, D., Lerach, S., Teubner, B., and Brindle, P. K. (2016) Mutation of the CH1 Domain in the Histone Acetyltransferase CREBBP Results in Autism-Relevant Behaviors in Mice, *PLoS ONE, 11*(1), e0146366.
22. Zurek, A., Kemp, S., Aga, Z., Walker, S., Milenkovic, M., Ramsey, A. J., Sibille, E., Scherer, S., and Orser, B. (2016) α5GABAA receptor deficiency causes autism-like behaviors, *Ann Clin Transl Neurol*, 3: 392–398.
23. Fatemi, S., Reutiman, T., Folsom, T., et al. (2010) mRNA and protein levels for GABAA α4, α5, β1 and GABABR1 receptors are altered in brains from subjects with autism, *J Autism Dev Disord,* 2010;40:743–750.
24. Fatemi, S., Reutiman, T., Folsom, T., Thuras, P. GABAA receptor downregulation in brains of subjects with autism. *J Autism Dev Disord* 2009;39:223–230.
25. Bruining, H., De Sonneville, L., Swaab, H., De Jonge, M., Kas, M., Van Engeland, H., … and Aziz, S. (2010) Dissecting the Clinical Heterogeneity of Autism Spectrum Disorders through Defined Genotypes (Genetic Subphenotypes in ASDs), *PLoS ONE*, 5(5), E10887.

Features

Soumyaa Mazumder '19

*The interview has been condensed in certain sections for clarity

This past September, THURJ had the opportunity to sit down Professor Matthew Meselson, Thomas Dudley Cabot Professor of the Natural Sciences in the Department of Molecular and Cellular Biology. Kind, charismatic, and generous with his time, Prof. Meselson spoke with us about his early beginnings in science, such as the famous 1958 experiment that proved the semiconservative replication of DNA, as well as about his current research and public policy interests. In addition to being a professor at Harvard and serving on the council of several eminent scientific councils, such as the National Academy of Sciences, Prof. Meselson is also the Co-Director of the Harvard-Sussex Program on Chemical and Biological Weapons.

SM: Thank you so much for your time today, Professor Meselson. To start off, one question I'm curious about is had you also been interested in science when you were younger? Like—

MM: [almost immediately] Yes.

SM: [laughs] Always. Why I'm asking is because I know you went to the University of Chicago and studied history and the classics there. And so I'm wondering whether that had originally been your main interest and if science came along later or had always been there.

MM: [laughs]. Yes, that was a detour. No, what happened was during the war, most young people didn't feel like just goofing off during summer vacation, so you either went and got a job or went to summer school. Each summer, I would do one or the other, and because I went to summer school, I got enough academic credits to get my high school diploma. So I went to the registrar at John Marshall High School in Los Angeles, and said "I've got all the credits I need, I'd like my high school diploma." And she said that California state law requires you have three full years of physical education. Well, I was very surprised by that. I didn't see what the relevance of that was, and I certainly didn't just want to hang around for another year and a half, taking gym! So, I started asking around to figure out what I could do, and someone told me, "Well, there's this place called the University of Chicago and they'll take you without a high school diploma." And I'd thought I'd go there and right away study chemistry, physics, and math.

And so that's what I did. I went there – I was sixteen years old – and it turns out there's no electives, except one course you could take, which was a philosophy course. Instead, we read the classics. And

I wouldn't have done that ordinarily had I gone on a regular path. I would have gone straight into some chemistry or physics concentration. But that was a wonderful thing; I'm very glad it happened.

SM: After your time at the University of Chicago, what did you study next?

MM: So I spent three years there. And then I spent a year doing nothing, just living in Europe. And then I went to Caltech and had to be a freshman all over again. But I did not like Caltech very much then, because except for Linus Pauling's course in elementary chemistry, there was a lot of memory involved, even in physics. So I left and went back to the University of Chicago and got what would have been a bachelor's degree in chemistry, except then they didn't give B.A.'s in chemistry. They gave me a letter that said we do not give a bachelor's degree in chemistry, but if we did, we would give one to Matthew Meselson. That letter enabled me to get into graduate school at Berkeley in physics. And I stayed there one year, and I didn't like it. It was huge. In those days after the war, they took many more people in physics than they were going to graduate with PhD's, and I knew I wasn't cut out to be one of those top-flight physicists. And besides, I wasn't interested in physics for physics; I was interested in applying it to biology.

So I was about to go back to the University of Chicago, because they had a program called mathematical biophysics. And those words sounded just right – math, bio, physics. But that one year I had been at Caltech, I'd done a research project for Linus Pauling in chemistry, and so I knew him and I knew his children. There was a party at his house one summer – that summer I had been planning on going back to the University of Chicago. And the party was at the swimming pool. I was in the water, and Linus Pauling, world's greatest chemist, said, "Well, Matt, what are you going to do next year?" And I said, "I'm going to go to the University of Chicago" And he looked surprised, and he said, "Why don't you come to Caltech and be my graduate student?" And I'm still in the water, so I just look up to him and say, "Well, ok!"

And that's how I got to be a graduate student at Caltech. I loved it there – it was wonderful. I did x-ray crystallography because I thought understanding the structure of large molecules would be important for understanding biology. And at that time, my major interest was to know how it is that you can put together ordinary atoms and create something that's alive.

SM: Do you find though that the studies you had done in history and the classics, even though you obviously have a career in science, that there were any lessons or experiences that still carry over in terms of scientific thinking or research?

MM: Yes, especially now. I remember reading the book, Chance and Necessity by Jacques Monod, a very famous Nobel Prize winning biologist. He and Professor Francois Jacob discovered how genes are regulated. And somewhere in this book, he says that it will never be possible to modify the human germline and that what we know about molecular biology already tells us of the impossibility of doing this. [finds the excerpt].

Yes, he writes, "Modern molecular genetics offers us no means whatsoever for acting upon the ancestral heritage so as to improve it with new features…on the contrary, it reveals the vanity of any such hope."

My major interest was to know how it is that you can put together ordinary atoms and create something that's alive.

Well, when I read that it was obvious that he was wrong. Because already by then, somebody named Beatrice Mintz had shown that if you take the embryo of a mouse and mix the cells with the embryo of a different mouse, you can get a mosaic mouse. And you can also get offspring from the mosaic mouse. And you can take a nucleus from one cell and put it into another cell. It was also known then that you can transform bacteria by changing the DNA. And so putting all these things together that were known at the time, it seemed that already then you could outline how you might be able to change the human germline by changing the DNA in a single human egg, re-implanting it into a uterus, and getting a baby that is genetically different in the germline. And I said in a talk I gave called "Biological Riddles to which the answer is man," this wouldn't happen for a while, but once it starts it will certainly become commonplace. For example, if you could tell people, "If you allow us to modify the germline of your offspring, they will never get cancer," and that turns out to be true, even people very much opposed to this kind of modification to the human germline will say "Well, wait it's very hard to say no to that."

Once it starts, it's very hard to stop. And it also raises the question, "What is it about being human that we really value?"

In fact, in that talk I previously mentioned to you, I specifically said, "Nevertheless, the prospect of human intervention into human nature, even if only theoretical for the time being, will cause us to be increasingly drawn to the study of human origins and the sources of civilization, in order to know what humanity is, in order to keep faith with it. Certainly the study of the humanities is an indispensable component of this endeavor."

So I think going to University of Chicago made me realize that there was this absolutely marvelous literature, particularly the Greek Classics and Homer, that put me into that frame of mind to understand these issues.

Recently as you probably know, the National Academy of Sciences has released a report on genetic engineering of the human genome.

In Chapter 5, they discuss changing the human genome in order to remove or change genes that cause bad diseases. That chapter recommends, that subject to certain guidelines and controls, that we go ahead with that. But the next chapter is called "Enhancing the Human Genome," and that's not to get rid of something bad. That's to, I don't know, do whatever you think will make people "better," and that the Academy says is too soon to start that. They don't say, though, don't ever do that. So this already shows an enormous shift in the way that people are thinking.

Once it starts, it's very hard to stop. And it also raises the question, "What is it about being human that we really value?"

We are a very transient form of Homo sapiens. You could say there will be worlds beyond worlds and men beyond men, big changes coming. H.G. Wells once wrote, humans could have organized society so that everyone would have enough to eat, shelter, a decent education, but our species hasn't chosen to do that. So, it's very hard to predict, but I suppose what it means to be human will change.

SM: Going back to the early beginnings of your career, you're very much one of the founding fathers of molecular biology in so many ways.

MM: [laughs] More like a cousin.

SM: You have many famous experiments for which you are known, such as showing the replication of DNA is semi-conservative, the existence of messenger RNA, restriction enzymes, etc. I just wonder, is there one experiment or topic in particular that has been the most exciting for you?

MM: Well of course it was the semi-conservative replication of DNA, for a number of reasons. First of all, because we had to do a lot of work to get there. We had to invent a new method for measuring densities of macromolecules. Second, when we finally got the result, it was all in a sudden flash in a dark room. You see one piece of film with a hybrid DNA band. Third, because we had such a pleasant working relationship, Frank and me. It was just marvelous. I was a graduate student living a wonderful graduate student life.

You mention mRNA. That was actually not my idea. That was the idea of Sydney Brenner. He came with Francois Jacob to my lab in Pasadena to do that particular experiment, partly because he was coming anyways to do a completely different experiment with me, which we therefore didn't do. And also because it was only in my lab that this experiment could be done. We had the necessary carbon and nitrogen isotopes and the necessary experience with cesium chloride density centrifugation experiments. So I worked on the centrifuge experiments, although not as much as Sydney did. It really wasn't my experiment; I was a collaborator.

However, I was doing another experiment at the time that really was my experiment, and that was looking into the basis of genetic recombination in lambda phage. There were two extreme ideas at the time. One was that recombination occurred by breakage and joining of 2 DNA molecules and you connect them up the other way around, and then you have recombinant molecules; so this requires breaking of 1 DNA molecule. The other idea was called copy-choice. That is you have two different DNA molecules, homologous, but not identical. And then you start copying off of one, and then you switch your copying machine (there is no such animal, but in those days we thought there could be), and then copy off the other DNA molecule. So now you have a molecule that has partly the sequence of 1 DNA and partly the sequence of the other.

And so at the time Sydney and Francois were doing their experiment, I had no family and obligations and I was doing this experiment on DNA recombination by myself, and I published the results together with John Weigle, a colleague in Switzerland. The excitement of seeing that result was not like seeing the DNA replication, but still great.

You could say there will be worlds beyond worlds and men beyond men, big changes coming.

And then there were some other things. I studied gene conversion in bacteria, and this made me wonder what good it is. I came to the view that maybe it is useful for repairing defects in DNA. How would that work? You would have to have some way in which the DNA-enzyme system would recognize a mismatch, a case in which the polymerase had made a mistake. And then it would go and remove this mismatch. But wait, if I remove either one base or the other with equal likelihood, this may perpetuate the mistake. For instance, let's say you have a sequence that should be A-T, and instead the polymerase has made a mistake and it is A-A. You should put a T in, but how do you know which strand to put it in? Well, I realized you needed a way to know which strand is the new one, because that's the strand where the mistake is. And I thought maybe that strand could be recognized by the methylation, since I knew that DNA does not get methylated right away after it is synthesized. It takes a little while and that would allow time for some kind of sensing enzyme to find a mismatch, a little bulge in the DNA, and then ask "Which of you two strands has methyl groups on it? Oh it's you, then I will chew away the stuff on the other strand and try again."

And so, I published the results of that experiment on mismatch repair. And there were a few other things.

SM: What research interests are you currently pursuing?

MM: What I'm interested in right now is the question of why does sex exist? I got interested in that because I was visiting with a friend at Yale, a famous old naturalist named Ian Hutchinson, and I asked him, why does sex exist? There were already at that time several explanations with no clear proof of which one is right. And he asked, "Do you know about the Bdelloid rotifers – these are little invertebrate animals, and no one has documented the existence of sexual reproduction or males." So we did a lot of work on these animals. At first we thought we were finding evidence that supported the possibility that they had evolved for many millions of years without any outcrossing. But now, because of DNA sequencing that we've been doing recently it's quite clear that they do engage in outcrossing. I'm quite interested in that now and I'm following it up. It turns out that our sequencing also indicates that they have a very unusual kind of meiosis, one that is only known in a certain kind of plants, which has implications for evolution that I won't go into.

And the other part of my life is involved in what you could call politics. More or less by accident I took a job just for the summer in 1963 at the State Department in the arms control and disarmament agency. They essentially let me do anything I wanted. I ended up looking at the possibility of arms control of biological weapons. I was interested in that because DNA engineering was just beginning to come along. Paul Berg and other people were splicing DNA and putting other pieces together. And here's my science, biology, used to wipe out all kinds of people. Not only that, but if you projected in the future, maybe you could even begin to affect the brain in warfare just opens up a lot of ominous things. So I decided to look into that, and I have spent half of my years at Harvard doing things about that.

What I'm interested in right now is the question of why does sex exist?

SM: Yes, I saw that you are involved with this program called the Harvard-Sussex program. Could you tell us a little bit more about your involvement with the program?

MM: Yes, I created that together along with a friend of mine called Julian Robinson. We've done a lot of things. We've published a journal for many years. But I also got involved with doing things on the ground, with combining science with arms control. Three things: the first was the American association for the Advancement of Science (AAAS) asked me to design a study of the ecological and health side effects of the herbicide spraying during Vietnam. I conducted a small pilot study for six weeks during the war in Vietnam. Here was a case, where this was not biological warfare in the usual sense but a form of chemical warfare (although it is against plants not people). Naturally I had some concern that you just don't want to get this kind of warfare started in any form, but I didn't go into there with the objective of stopping it, just studying it.

But I learned a lot of things from that. And eventually we in fact did get the whole program stopped. In particular, the last day I was there, the commanding general asked me to come to his office and tell him what I thought about the military utility of the herbicides. I said I have no idea about that. And he said, "You what to know what I think?" And I said, "Yes sir." He said, "I think they're shit." And I said something like, "But General Abrams, you're the commanding officer of all the American forces. Why are we using it?" And I distinctly remember what he said next. He said, "You don't understand anything about this war young man." This was in 1970. "He said, those decisions are made in Washington."

I wanted to be sure my memory of this was exact. And I remembered his son, John, who was a captain in Vietnam agreed with him. So I found John, himself retired as a four-star general who now runs a consulting company down in Washington, and I asked him, "Is that right? Is that what your dad thought? Is that what you thought?" [laughs] He sent me back an email which is much stronger than what General Abrams said.

This experience showed me one of the really great difficulties of trying to understand what war is all about. On one hand the joint chiefs and a lot of people were saying you're helping us to win the war (which we of course did not win), and on the other hand people like the commanding general were saying this type of warfare is useless; and the war, it still goes on.

SM: Do you have continued involvement in Washington? Are there any studies that are continuing with regards to current wars or biological warfare?

MM: Yes, lots. President Nixon renounced biological weapons categorically, and I had a lot to do with that. I wrote many papers for Henry Kissinger. Henry used to have his office in the building next door to the Semitic Museum. There was on the third floor a kitchen, and two Hungarian ladies made lunch every day. That's how I got to meet him. He also participated in a Harvard-MIT arms control seminar that met in evenings, and I used to go to that. So when he was appointed as National Security Advisor by President Nixon he asked me to write him a paper (because by then he know I was interested in biological weapons). So I wrote him some papers. Basically I made the argument that biological weapons are really cheap. Which they are. And they can wipe out lots of people. Which they can. We have nuclear weapons. Do we really want to pioneer the introduction into the world of a weapon of very great mass destructive capability

that everyone can have? Or is it best if it so expensive that no one can afford it? The next best is that only the US could afford it. And so on. And that was a winning argument.

I believe that people in the science field shouldn't just be incredibly specialized. They should know something about all aspects of science.

Before that, no one really put the argument that way except a few pals who I got to know in the Defense Department. Instead they argued, we better do this because someone else is doing it. Or maybe we better do it to find what the potential is out there. But no one was saying, wait that's a door best left closed. If we develop it, how can we go on a campaign to get others to stop or not start? For example, when I was working at the state department, one of the things I did after deciding to work on the problem of biological weapons was to go to Fort Detrick, where we were then developing biological weapons. I was given a tour by a very nice young man named LeRoy Fothergill, a fairly eminent biologist. He had been on the faculty at Harvard Medical School before he had been joined the army. We came to a building, and it seemed like an ordinary building about seven stories high that looked like it had windows. But when you got up close, you saw that it had phony windows. I asked him what is this, and he said, "It's a big fermenter, we make anthrax there." And I said, "Why do we need to make anthrax?" to which he responded, "Well, it will save money. It's a lot cheaper than nuclear weapons."

That's the point when I suddenly realized, "Wait, we don't want a weapon that's a lot cheaper." I went back to my office. My office mate was Freeman Dyson, the physicist. I had taken quantum mechanics from freeman at Berkeley, so I knew him. And he was very encouraging. I remember him saying that my intuition about this is wiser than I knew. Freeman was always someone who thought developing biology for hostile purposes was really a dangerous thing to do. And so I spent years trying to persuade people to get out of this business, and eventually President Nixon did it.

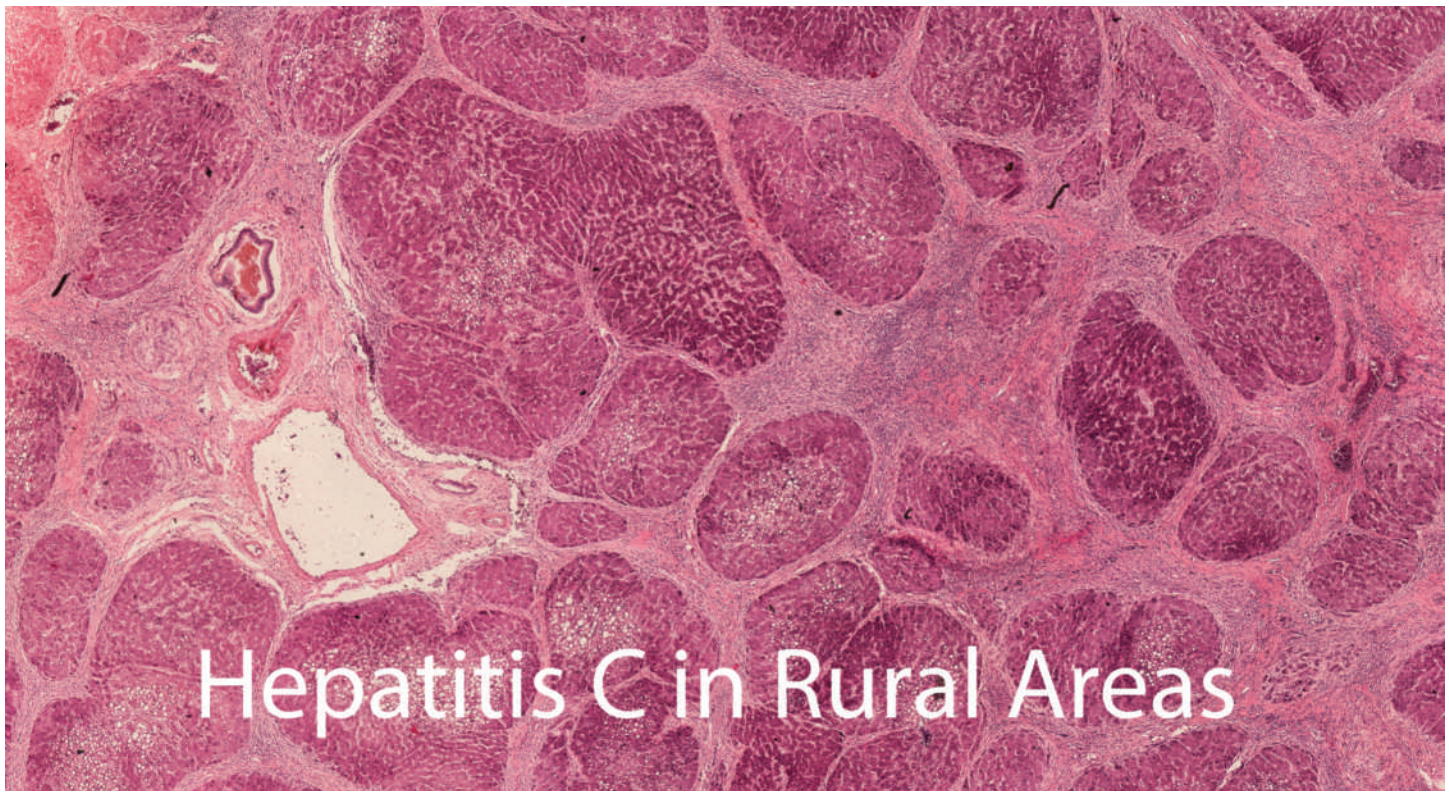
SM: It must be so satisfying to not only conduct exciting basic science research in your lab, but to also be part of these bigger policy changes in Washington. Was it gratifying to give these reports to generals and higher-up officials in Washington and know the impact that these contributions would make?

MM: For me, it's always been very much combining science with some arms control objective. In a sense there's only really one science. We have a National Academy of Sciences, plural, but that's not right. It should be a National Academy of Science, singular. Because I believe that people in the science field shouldn't just be incredibly specialized. They should know something about all aspects of science. So it's really very pleasant to be able to change.

The basic idea I have always had in my mind was to prevent our science, biology, for terrible purposes. People go blast each other and slice each other up. Every big technology that humans have ever devised – fire, gunpowder, nuclear fission, electronics – has been exploited for peaceful purposes as well as military purposes. The one exception is biology, molecular biology. That has the possibility not just to kill people but of changing them ultimately, not only by changing what it means to be human, but also by dissolving the difference between war time and peace time. And maybe because as any biologist would feel, this is my science. I don't want my science to be used for hostile purposes.

SM: Thank you so much for your time.

MM: You're welcome, thank you.



Hepatitis C in Rural Areas

Dalton Brunson '19

Author Note:

This literature review draft was prepared for the National Patient Advocate Foundation, a non-profit organization who works in conjunction with the Patient Advocate Foundation. A special thank you to Marguerite Beiser, NP of the Boston Health Care for the Homeless Program for providing insight into state specific Medicaid information. Initial literature search was completed by Anitra Persaud, a fellow at the National Institutes of Health.

BACKGROUND

Hepatitis C (HCV), which has come to be known as a silent killer, affects an estimated 3.2 to 4 million people in the United States alone with an estimated 25-50% of cases being undiagnosed [1,2]. HCV caused more deaths in 2013 than sixty other communicable diseases combined including other well-known infections such as human immunodeficiency virus (HIV), tuberculosis, and pneumococcal disease [3].

Like HIV, HCV is easily and primarily transmitted through percutaneous exposure to blood [4]. With blood as an infection vector, the "baby boom" generation (born 1945-1965) is primarily at risk for HCV due to unsafe medical procedures and unscreened blood transfusions. However, scientific breakthroughs in testing and improved blood supply screening have led to a decrease in transmission through infected blood supplies and medical procedures [5]. While the incidence of new cases has declined in the older "baby boom" generation, educational materials published by the Centers for Disease Control and Prevention (CDC) state that baby boomers are more than five times as likely to have HCV when compared to other adults [6]. The top HCV exposure is now related to illicit drug use, which accounts for roughly 60% of new cases in the United States [7]. Interestingly, HCV infection profiles change with different cultures as well as access to medical and screening procedures changes the demographic of the infected population from country to country [5].

While the disease is not immediately life threatening and is generally asymptomatic, HCV can have many adverse effects on an infected individual's health [8]. Of the people who are infected, the disease may spontaneously clear in approximately 20% of cases. The exact biological mechanisms behind the ability to spontaneously clear HCV are currently unknown [1]. In the other cases, HCV will progress to a chronic stage. Patients with chronic HCV face an increased risk from a litany of complications that include increased risk of bleeding, liver cancer and even death [9]. Untreated chronic hepatitis C often leads to end-stage liver disease and requires a liver transplant [7]. Unsurprisingly, sufferers of untreated chronic hepatitis C can expect shortened life spans [8].

TREATMENT

Diagnosis

The CDC has outlined a series of recommendations on frequency of testing for individuals at-risk for contracting HCV. Current CDC

guidelines recommend that anyone born between 1945-1965 be tested at least once for HCV [4]. More rigorous testing schedules are recommended for people who inject drugs and HIV-seropositive men who have unprotected sex with men. Data, aligned with current recommendations, suggest that the incidence of infection is largest among people who inject drugs. Currently, the process for diagnosing HCV is a multistep (multi-visit) process that requires several laboratory diagnostic tests.

Even as diagnosis tools have improved, researchers estimate that up to half of people living with HCV are living undiagnosed [1]. In the past, campaigns have relied on passive testing or only testing those who reported risk factors; however, many of the behaviors associated with HCV transmission are considered socially taboo, leading individuals to under report such behaviors. Attempting to capture these underreported populations, by utilizing "opt-out" testing, Galbraith et al. (2015) adapted the method used to test the baby boom population at emergency room departments. The researchers integrated HCV antibody testing into the standard clinical care for anyone who came to the emergency department who did not choose to decline the test. The study concluded that there was high prevalence level of unknown HCV in emergency departments that would not have normally been diagnosed or recognized [10]. In total, the researchers found a prevalence rate three times higher in the "opt-out" screened population in the emergency department than the estimated general population prevalence. This spike may be explained by the type of patient who utilizes emergency departments. However, it is important to note that without integrating the HCV testing into the standard of care, many of these infections would have continued to be undiagnosed.

Historical Treatments

Sufferers of HCV should be able to find relief in that this condition can be considered a treatable disease. Successful treatment of HCV is termed a sustained viral response (SVR), indicated by an undetectable HCV RNA level [11]. Those patients who undergo treatment and achieve SVR can expect a life span similar to those without HCV [8]. Treatment of HCV has continuously evolved since the diseases discovery in 1989 [5]. The first standard of care was a combination therapy of peginterferon alfa (PEG) and ribavirin (RBV), which led to SVR rate of approximately 50% [11]. Unfortunately, SVR was not achieved at rates across all ethnic populations. While minorities have a history of being underrepresented in clinical research, a review by Lisker-Malman & Walewski (2013) contained almost identical sample sizes of Caucasian and African American patients, and the African American patients achieved a SVR at rates much lower than their Caucasian counterparts (52% vs. 28%). The same disparity existed in other studies between Caucasians and Latinos. The researchers were careful to explain that some of the difference in treatment response was more than likely due to social, economic, and co-morbidities. However, the most interesting finding pertains to single nucleotide polymorphisms (SNP). In the review, the team found that patients with a particular SNP were more likely to achieve SVR. The data was supported in that minorities with the same genetic SNP achieved SVR at the same rates as their Caucasian counterparts. The PEG/RBV therapy works by activating the innate immune system, but such a treatment program favors certain SNP variants that predominate in certain ethnicities due to global dispersal [12].

In 2011, a new class of drugs called direct acting antivirals (DAAs) was introduced to treat HCV. The DDAs act directly on the virus, instead of relying on the innate immune system. The introduction and future development of additional DAAs will eventually eliminate the disparities faced by those who possess SNPs associated with lower abilities to achieve SVR on the PEG/RBV combination therapy [1] . The first generation of DAAs introduced to the market was telaprevir and boceprevirl which were recommended to be used in combination with the standard PEG/RBV therapy. The first generation DAAs had lengthy treatment plans that lasted a minimum of 24 weeks. In addition to the lengthy treatment regimens, the use of a PEG/RBV therapy is associated with adverse side effects such as impairment in concentration, depression, insomnia, and irritability [13]. Greebly et al. (2007) attempted to ascertain if directly observed therapy (DOT) would improve rates at which of patients who inject drugs were able to achieve SVR. The relevant finding are that even with DOT a considerable percentage of the cohort dropped out of the study due to the adverse side-effects of the PEG/RBV therapy [14]. Similar to the PEG/RBV therapy, the first generation DAAs were found to have increased adverse side effects and had to be used in combination with the PEG/RBV therapy, but they were also able to improve SVR rates to 63-75% in certain strains of HCV [5].

Current Treatments

HCV has many biological characteristics, that when combined, make HCV difficult to treat. The virus has a high rate of replication with no DNA replication proofreading mechanisms, which leads to increased genetic variability. The increased genetic variance has led to several different strains of HCV to emerge across the globe. Astonishingly, different variants of HCV can have genetic profiles that vary by as much as 30% [5]. While Genotype 1 accounts for the majority of HCV cases, the prevalence of different strains is not evenly spread across the globe with different strains being concentrated in different populations. Such diversity not only presents challenges to developing effective treatments but also creates challenges for generating accurate epidemiological data [5].

The advent and the subsequent 2013 FDA approval of second generation DAAs made the possibility of treating multiple strains of HCV a reality [7,15]. To reflect the emerging personalized medicine in HCV treatments the American Society for the Study of Liver Disease (AASLD) and the Infectious Diseases Society of America

(IDSA) then recommended that an assay be performed to determine the genotype of the HCV variant prior to treatment. They consider the genotype recommendation to be a Class I, Level A (the highest possible) recommendation within their prescribed set of guidelines meaning that the genotype determination is now paramount for successful treatment and best outcomes [4].

To attest to the astounding rate at which HCV treatment is evolving, the guidelines published by the Federal Bureau of Prisons recommend consulting the AASLD website before treatment begins. Furthermore, a list of the most common standard of care DAAs is outlined with their potential uses. When looking at the long list, each drug must be used in combination with another to prevent creating resistant strains, and each drug is only approved for treatments of certain genotypes of HCV [5,16]. Moreover, the second-generation DAAs dramatically improved patients' abilities to achieve SVR. Many clinical trials suggested that the second generation combination therapies were able to achieve SVR in upwards of 90% of patients, and the new DAAs had almost no side-effects, and they found that patients tolerated the new drugs almost as well as the placebo in the control arm of studies [8]. While second-generation DAAs are a miraculous breakthrough that could lead to a tremendous reduction of the prevalence of HCV, there are many barriers to treatment that patients and possibly providers face.

Cost

Undeniably, the most well documented barrier to accessing current standard of care DAAs for HCV treatment are their exorbitant costs. Initially, when second generation DAAs were introduced to the market, some were priced at over $1,000 a pill [17]. It is troubling to know that if the whole-sale acquisition price were paid for Sovoldi, a popular second-generation DAA, to treat everyone in America with HCV, the cost would easily have exceeded the total cost for all prescription medications in 2011 [7]. While it should be noted that drug companies are for-profit organizations that spent enormous sums of money to develop new drugs, the profit ratio that Gilead experienced on the drug Sovoldi was 20:1, a ratio that a Congressional panel determined to be excessive [18].

However, while the costs of second generation DAAs such as Sovoldi and Harvoni are large in comparison to the costs for treating other disease, a recent review of the pricing and affordability of DAAs for HCV treatment did observe that the total cost of treatment for DAAs was roughly equivalent to the original PEG/RBV therapies [17]. Therefore, the treatment would be money saving over-time due to the increased cure rates. Unfortunately, the upfront cost of the therapy is still prohibitive of wide scale utilization, and it eliminates the possibility for patients to utilize older more-cost effective drugs since the new drugs are already more cost-effective, albeit with a much larger up-front cost [17,19].

Additionally, many government agencies have worked together to increase buying power in an attempt to secure discounts from the wholesale acquisition price [3]. However, such claims are difficult to verify because of a continued lack of transparency from pharmaceutical companies as they consider the actual price that a consumer pays to be part of a private business contract. Such lack of transparency has led to a great disparity within different organizations ability to afford DAAs [17]. Some claims in various reports show that government agencies such as the Department of Veterans Affairs and the Department of Defense receive discounts of roughly 24% while other agencies such as Medicare and Medicaid receive discounts estimated at 50%, relative to the wholesale acquisition price [3]. Nonetheless, such discounts are not enough to ensure equitable access to all people who are diagnosed with HCV and are recommended to receive DAAs as the standard of care.

Perhaps one main cause of the delay in change is the substantial amount of literature that concludes treating individuals who have F1 and F2 stage liver disease is not cost effective. A review complied by Rosenthal and Graham (2016) shows ample literature that suggests the most cost effective method for treating those with HCV is to prioritize treatment for the youngest individuals with advanced fibrosis [17]. The researchers conclude that it would be more cost effective to wait until liver disease had progressed to moderate to advanced stages; however, the models that make such claims fail to account for factors other than direct medical costs and omitting factors such as treatment or prevention. A person who is at high risk for spreading the disease, such as person who injects drugs or a member of the prison population, would be less likely to spread the disease if treatment were offered in the early stages of the disease. Additionally, the models fail to account for quality of life and other major contributing factors [17]. According to the AASLD, once a person experiences advanced stage liver disease (F3-F4), the risk of developing complications such heptocarcinoma (HCC) is considerable. Furthermore, individuals who have advanced fibrosis remain at risk for HCC even after SVR is achieved, thereby reducing their lifespan [20]. However, another cost analysis study that took more variables into consideration concluded that DAAs should be given to every eligible person waiting on treatment [21].

Coverage and Access

With the exorbitant cost of DAAs, many insurers, especially government-based programs such as Medicare and Medicaid, require patients to go through an arduous pre-approval process before they can receive access to DAAs. In efforts to curtail costs, insurance companies have created various requirements that patients must satisfy before becoming eligible for DAA treatment. The requirements vary greatly from state to state, which can affect the ability to translate research from the pre-approvals process from one state to another. The four main categories from which each state draws their restrictions the progression of liver fibrosis, patient absence

from alcohol and drug use, HIV co-infection status, and prescriber limitations. The restrictions contradict the guidelines outlined by the AASLD and IDSA, which may lead one to draw the conclusion that the restrictions associated with the pre-approvals process are only in place to restrict access to expensive medications and create disparities in health [22]. Ultimately, the pre-approvals process, a by-product of the large cost of DAAs, serves an institutional barrier of care for some of the most vulnerable patients with HCV.

One study found that insurance type was a contributing factor in the ability to access DAA treatment for HCV. After analyzing a large cohort of patients from multiple states, it was found that one in six patients were outright denied access to DAA therapy. Unsurprisingly those with government-based insurances were more commonly denied treatment, with Medicaid having the greatest number of denials [2]. The Lo Re et al. (2016) study researched disparities by insurance type and concluded that almost half of patients who were on Medicaid, the most vulnerable HCV patients, were outright denied treatment for DAAs [2]. Of the patients with advanced cirrhosis, one of four individuals with Medicaid were denied treatment compared to almost no denials of comparable patients on other forms of insurance, including the government run Medicare. In addition, patients who were allowed to access DAAs on Medicaid were required to wait longer for treatment.

Another study completed by Do et al. (2015), found that only one in ten patients were denied treatments, and they even drew the conclusion that Medicare and Medicare were significant predictors of approval and decreased wait times on appeal decisions with a higher percentage of appeals approved, in stark contrast to the study published by Lo Re and colleagues [23]. Even Do recognizes that his participants' abilities to access care might be an overestimate because his subjects were in the care of specialists at major tertiary hospitals with liver clinics. One can conclude that the ability of patients to access DAAs, even when on Medicare and Medicaid, can be greatly improved based on healthcare setting; however, one can also observe that one in ten patients are still not getting care even when accessing the highest level specialty providers.

The conflicting information leads to the need for further research into why patients on Medicaid are disproportionately not able to access DAAs for HCV treatment. Do et al. (2015) clearly showed that patients on Medicaid are able to access treatment when given specialized care, but further research needs to be conducted to determine why a larger cohort of Medicaid patients who are not privy to such advanced care are unable to receive treatment. Another consideration is the relatively relaxed preauthorization requirements in Connecticut (the location of the Do study).

Current and Future Directions

Cost and cost effectiveness is a direct diving force in determining the various payers' decision to cover DAAs, but it is almost impossible to know the cost at which individual buyers acquire DAAs because pharmaceutical companies consider the exact prices as confidential business information. The cost of DAAs across the globe varies widely showing increased variability among high-income countries without regard to the countries annual gross national income. With such variability among price in high income countries, countries should observe the world stage to determine the best practices for obtaining the most affordable DAAs for their citizens. [21].

The United States also faces another barrier to lowering cost in that it does not have a way to increase its buying power. The states that participated in pooled buying with other agencies within their state were able to procure DAAs at lower costs. The departments that did not practice any cost lowering measures pay, which only confirms that increased purchasing power will only serve to decrease the cost of DAAs [3]. The Massachusetts Department of Health and Human Services further showcased the ability of increased purchasing power to lower the cost of prescription drugs when it announced in June 2016 that it had reached a deal with Gilead "significantly" reduce the price of Harvoni [24].

Unfortunately, increased buying power can only decrease costs to a limited extent. Some suggest that the federal government take a drastic step of non-voluntarily seizing the patent for DAAs based on eminent domain with the claim that cost to manufacture would be reduced to just $3 a pill. Such a tactic is not without precedent. The government seized the patent for the antidote to anthrax following the events of September 11th. However, that drug manufacture lowered the price of the drug to retain the patent [7].

While some states have been proactive in removing certain sanctions, the removal has no uniformity from state to state. For example, the restriction lifted at Tufts only removed the need for advanced liver disease [25]. The change did nothing to eliminate the requirement for abstaining from alcohol or drugs; failing to realize a growing body of literature that suggests that treatment can in-fact be a means of prevention and cost reduction. New York has made some of the most sweeping changes when the Attorney General announced in April 2016 that nearly all commercial health insurance plans in the state would remove the requirements for liver fibrosis, the use of drugs, or restrict the authorizing physician to be a specialist [26]. Unfortunately, even in states where institutional barriers, such as preauthorization, have been lessened to improve access to DAAs, only 14% of those who are diagnosed with active HCV are in treatment [20].

CONCLUSION

HCV is a silent epidemic that affects a considerable portion of the world's population. Fortunately, the disease is considered treatable. In light of the prohibitive costs associated with HCV treatments, it is imperative that government agencies pool together resources to increase purchasing power and practice cost saving measures to

further expand care for HCV patients. While legal action is reducing some barriers to care at the institutional level, many barriers still exist for those insured by Medicaid. More research should be undertaken to discover what factors from the patient and provider perspective are most prohibitive to seeking DAAs for HCV treatment in an effort to further eliminate institutional barriers to care.

While barriers exist for those who are currently seeking treatment, large portions of infected individuals are not seeking testing or treatment. Future studies on patient perspectives are imperative for understanding why patients are not seeking care in less restricted settings. Furthermore, no current studies examine how current policies lead to increased racial and ethnic disparities in HCV treatment. Combining research with knowledge already know about HCV and its treatment options, public health organizations will be able to cater individualized campaigns to patients, providers will be able to reduce treatment nonadherence and improve access, and government agencies will be able to see what specific regulations are the most prohibitive to care. All of which are necessary if we hope to lessen the disease burden associated with the domestic and global HCV endemic.

References

[1] Lisker-Melman, M., & Walewski, J. L. (2013). The impact of ethnicity on hepatitis C virus treatment decisions and outcomes. Dig Dis Sci, 58(3), 621-629. doi:10.1007/s10620-012-2392-0

[2] Lo Re, V., 3rd, Gowda, C., Urick, P. N., Halladay, J. T., Binkley, A., Carbonari, D. M., . . . Kostman, J. R. (2016). Disparities in Absolute Denial of Modern Hepatitis C Therapy by Type of Insurance. Clin Gastroenterol Hepatol, 14(7), 1035-1043. doi:10.1016/j.cgh.2016.03.040

[3] Beckman, A. L., Bilinski, A., Boyko, R., Camp, G. M., Wall, A. T., Lim, J. K., . . . Gonsalves, G. S. (2016). New Hepatitis C Drugs Are Very Costly And Unavailable To Many State Prisoners. Health Aff (Millwood), 35(10), 1893-1901. doi:10.1377/hlthaff.2016.0296

[4] American Association for the Study of Liver Diseases; Infectious Diseases Society of America (AASLD; IDSA). Recommendations for Testing, Managing, and Treating Hepatitis C [Internet]. 2016 [cited 2016 Aug 17]; Available from: http://hcvguidelines.org/sites/default/filesA/HCV-Guidance_July_2016_b.pdf.

[5] Ansaldi, F., Orsi, A., Sticchi, L., Bruzzone, B., & Icardi, G. (2014). Hepatitis C virus in the new era: perspectives in epidemiology, prevention, diagnostics and predictors of response to therapy. World J Gastroenterol, 20(29), 9633-9652. doi:10.3748/wjg.v20.i29.9633

[6] Centers for Disease Control and Prevention. Hepatitis C: Why Baby Boomers Should Get Tested [Internet]. [updated 2015; cited 2016 Aug 16]; Available from: http://www.cdc.gov/knowmorehepatitis/Media/PDFs/FactSheet-Boomers.pdf.

[7] Trooskin, S. B., Reynolds, H., & Kostman, J. R. (2015). Access to Costly New Hepatitis C Drugs: Medicine, Money, and Advocacy. Clin Infect Dis, 61(12), 1825-1830. doi:10.1093/cid/civ677

[8] Leleu, H., Blachier, M., & Rosa, I. (2015). Cost-effectiveness of sofosbuvir in the treatment of patients with hepatitis C. J Viral Hepat, 22(4), 376-383. doi:10.1111/jvh.12311

[9] Food and Drug Administration. FDA Approves Epclusa for Treatment of Chronic Hepatitis C Virus Infection [Internet]. 2016 [updated 2016 Jun 28; cited 2016 Aug 17]; Available from: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm508915.htm.

[10] Galbraith, J. W., Franco, R. A., Donnelly, J. P., Rodgers, J. B., Morgan, J. M., Viles, A. F., . . . Wang, H. E. (2015). Unrecognized chronic hepatitis C virus infection among baby boomers in the emergency department. Hepatology, 61(3), 776-782. doi:10.1002/hep.27410

[11] Le, T. K., Kalsekar, A., Macaulay, D., Yuan, Y., Sorg, R. A., Behrer, C. R., . . . Wu, E. Q. (2015). Treatment patterns, health care resource utilization, and costs in U.S. patients diagnosed with chronic hepatitis C infection who received telaprevir or boceprevir. J Manag Care Spec Pharm, 21(4), 308-318. doi:10.18553/jmcp.2015.21.4.308

[12] Race, E., & Genetics Working, G. (2005). The use of racial, ethnic, and ancestral categories in human genetics research. Am J Hum Genet, 77(4), 519-532. doi:10.1086/491747

[13] Harris, M., & Rhodes, T. (2013). Hepatitis C treatment access and uptake for people who inject drugs: a review mapping the role of social factors. Harm Reduct J, 10, 7. doi:10.1186/1477-7517-10-7

[14] Grebely, J., Raffa, J. D., Meagher, C., Duncan, F., Genoway, K. A., Khara, M., . . . Conway, B. (2007). Directly observed therapy for the treatment of hepatitis C virus infection in current and former injection drug users. J Gastroenterol Hepatol, 22(9), 1519-1525. doi:10.1111/j.1440-1746.2007.05032.x

[15] Gilead. U.S. Food and Drug Administration Approves Gilead's Sovaldi (Sofosbuvir) for the Treatment of Chronic Hepatitis C [Internet]. 2013 Dec 2013 [cited 11 Jan 2017; Available from: http://www.gilead.com/news/press-releases/2013/12/us-food-and-drug-administration-approves-gileads-sovaldi-sofosbuvir-for-the-treatment-of-chronic-hepatitis-c

[16] Federal Bureau of Prisons. Evaluation and Management of Chronic Hepatitis C Virus (HCV) Infection [Internet]. 2016 Apr [cited 2016 Aug 17];]. Available from: https://www.bop.gov/resources/pdfs/hepatitis_c.pdf.

[17] Rosenthal, E. S., & Graham, C. S. (2016). Price and affordability of direct-acting antiviral regimens for hepatitis C virus in the United States. Infect Agent Cancer, 11, 24. doi:10.1186/s13027-016-0071-z

[18] Hatch, O.G., Wyden, R. United States Senate Committee on Finance. (2015) The price of Sovaldi and its impact on the U.S. health care system (97-329-PDF). Washington, D.C.: U.S. Government Publishing Office.

[19] Hoofnagle, J. H., & Sherker, A. H. (2014). Therapy for

hepatitis C--the costs of success. N Engl J Med, 370(16), 1552-1553. doi:10.1056/NEJMe1401508

[20] Barua, S., Greenwald, R., Grebely, J., Dore, G. J., Swan, T., & Taylor, L. E. (2015). Restrictions for Medicaid Reimbursement of Sofosbuvir for the Treatment of Hepatitis C Virus Infection in the United States. Ann Intern Med, 163(3), 215-223. doi:10.7326/M15-0406

[21] Andrieux-Meyer, I., Cohn, J., de Araújo, E. S. A., & Hamid, S. S. (2015). Disparity in market prices for hepatitis C virus direct-acting drugs. The Lancet Global Health, 3(11), e676-e677. doi:10.1016/s2214-109x(15)00156-4

[22] Center for Health Law and Policy Innovation of Harvard Law School (CHLPI). Examining hepatitis C virus treatment access: A review of select state Medicaid fee-for-service and managed care programs. Cambridge, MA: CHLPI

[23] Do, A., Mittal, Y., Liapakis, A., Cohen, E., Chau, H., Bertuccio, C., . . . Lim, J. K. (2015). Drug Authorization for Sofosbuvir/Ledipasvir (Harvoni) for Chronic HCV Infection in a Real-World Cohort: A New Barrier in the HCV Care Cascade. PLoS One, 10(8), e0135645. doi:10.1371/journal.pone.0135645

[24] MassHealth; Executive Office of Health and Human Services. MassHealth Implements New Drug Rebate Program, Expands Access to Hep C Treatment [Internet]. 2016 Jun 20 [cited 2016 Aug 17]; Available from: http://www.mass.gov/eohhs/gov/newsroom/press-releases/eohhs/masshealth-implements-new-drug-rebate-program.html.

[25] Freyer F. Tufts Health Plan lifts restrictions on hepatitis C drugs. Boston Globe [serial on the Internet]. 2016 Apr 20 [cited 2016 Aug 17]: Available from: https://www.bostonglobe.com/metro/2016/04/19/tufts-health-plan-lifts-restrictions-hepatitis-drugs/2EdZqgl8nXArTmA74fWOLN/story.html.

[26] New York State Office of the Attorney General. A.G. Schneiderman Announces Major Agreement With Seven Insurers To Expand Coverage Of Chronic Hepatitis C Treatment For Nearly All Commercial Health Insurance Plans Across New York State [Internet]. 2016 [updated 26 Apr 2016; cited 2016 Aug 18]; Available from: http://www.ag.ny.gov/press-release/ag-schneiderman-announces-major-agreement-seven-insurers-expand-coverage-chronic.

Daniel Kim '19

DK: I wanted to start out by talking a little bit about your background, your story: your research, but also how you grew up, where you came from?

LG: I grew up in Brookline, and my dad was a professor of orthopedic surgery at Harvard Medical School. He shared the Department of Orthopedic Surgery at MGH and then at Children's. But he, like me, was a physician-scientist, and he was really a biophysicist and a biochemist. He basically founded the orthopedic research labs at both places. So he was certainly an influence in my life.

I went for high school to the Winsor School. And then I went to Harvard undergrad and Harvard Medical School. At HMS, the immunology block was at the end of the first year and that's when I fell in love with immunology. That was it. I spent my fourth year at medical school in a laboratory here at Dana-Farber – Harvey Cantor's laboratory – where I learned about immunology, did immunology research, and then I trained at Massachusetts General Hospital in Internal Medicine followed by subspecialization in Rheumatology. After a postdoctoral fellowship at NIH in immunology, I returned to Harvard to become an Assistant Professor and start my own laboratory, in the Longwood Medical Area. My lab was at the Harvard School of Public Health with a joint appointment at Harvard Medical School where I rose through the ranks, and received tenure at the School of Public Health and the Medical School in 1990.

Immunology is something I've always loved, but I've also drifted into different fields, I've gone where the science takes me.

Basically, I have spent my life in Boston. At some point, I decided that it would be interesting to live somewhere else and do something different. When I was asked if I would consider being the Dean at Weill Cornell, I thought I should look at this opportunity, because I could perhaps have a greater impact by enabling other people's science and other people's careers if I were Dean, rather than just focusing on my own laboratory. And I just thought it would be fun to have two different careers in one's life. My first career was primarily as a basic researcher, also as a physician, although I didn't spend a lot of time in patient care. And I've been very fortunate to have a rewarding scientific career – far better than anything I would have ever dreamt I could have – so I figured it was time to give back. I had a fairly large research lab. When I moved to Weill Cornell, it got smaller. I decided to focus on just a couple of things. That's how I ended up at Weill Cornell. Which was great – I had a great time there, it was a huge job, much of which I was not prepared for, so there was a very steep learning curve which I always like. My intent was to stay at Weill Cornell another five years, but when Dana-Farber

asked me if I would consider the position of CEO and President, I just couldn't resist because it's such a fantastic institution. And it brought me back to Harvard, to Boston. Which was always my intention; I was always going to come back here at some point. Two of my three kids are here; my older son went to Harvard Medical School. He's a cardiothoracic surgeon at MGH. My younger son went to Harvard undergrad and thence to the Marine Corps where he was a Captain in the Marine Corps in Force Recon. , He then went to MIT business school, and now he's in business and politics. So two of my three kids are here and my daughter is in Washington, D.C. So there were a lot of reasons to come back. And I think Boston is an amazing place for the life sciences. It clearly still is the premiere place for the biomedical sciences. And it's not just that it has a plethora of top-notch academic medical centers, and research institutes, but it also of course has the whole venture capital biotech pharma environment which works nicely with academia.

DK: You're considered a pioneer, a champion of the field of immunology. Is there any direction you see the field going, either in cancer therapy or implications for other fields that really excites you? Or, could you tell us a little bit about what that search was like in the past, in your previous research – finding markers like Nk1.1?

LG: (laughing) That was so long ago!

But I'll tell you what my philosophy is. And I decided this when I started out: that I did not want to do humdrum science. I wanted to have the chance to make big contributions and the only way you can do that is to be a risk taker. And I kind of have a risk-taker phenotype anyway, so I decided I would just aim very high, and if I crashed, I crashed! Being a physician opens lots of avenues for various careers.

If you're not failing in some projects, you're not taking enough risks.

I decided that when I set up my own lab that I would try to figure out the relationship between the major histocompatibility complex and how that related to activating T cells by the T cell receptor. Then I became interested in gene regulation, which has really been my love, and wanted to figure out why an uncommitted progenitor cell, a naïve, T lymphocyte – decided to become one kind of T cell or another kind of T cell. And so we isolated several of the transcription factors that really controlled lineage commitment. Whether a T cell became a type 1 T cell, type 2 T cell, helper cell or killer cell. Probably what I'm best known for is identifying the master regulator transcription factor for type 1 immunity. This was T-bet. And then we also discovered the first transcription factor that controlled a key step in B lymphocyte development- from a B cell into an antibody-secreting plasma cell. And that led us into the ER stress response… We kind of go where the science takes us, and we've been fortunate to make some major contributions.

You can not be superficial, not just make a discovery and then leave it but pursue it – get down into the weeds.

To be a scientist, you have to tolerate highs and lows. The highs are great, when you make a big discovery; but there are a lot of lows. It requires emotional stamina. Everybody's experienced lows. Everyone is going fail at times. If you're not failing in some projects, you're not taking enough risks. That's the bottom line. So you have to fail at some projects. But if you're lucky, you'll succeed in a few, and that's all you can hope for. And we were pretty lucky.

Also, it's good to have an open mind. If you get an unexpected result, don't just ignore it. We isolated a gene called Shnurri-3 that we thought might be important in the immune system. But it turned out to not affect the immune system but instead had an enormous effect on the skeletal system. That led to a sustained interest in skeletal biology, and identification of new proteins that control bone mass. Osteoporosis (low bone mass) is a very common disease and we don't have ideal treatments for it. So, one goes where the science takes you; at the same time, it's important to focus on two or three different areas so that after the initial discovery one can pursue it in depth.

So where do I think cancer [therapy] is going? It truly is an amazing time for cancer. It really is. Two revolutions have occurred in the last decade – precision or targeted medicine which identifies unique genetic mutations in each patient's tumor and immunotherapy which has finally seen some astonishing successes for some tumors. My own lab has focused on the role of endoplasmic reticulum (ER) stress both in the tumor and also in the immune system with some interesting results. We predicted that if you manipulated ER stress pathways in the tumor cell, it should lead to increased tumor death. And we were right about that. What we didn't predict was that if you block the pathway in the immune system, you would actually activate antitumor immunity in multiple immune cells. It's like a double whammy, which is why we founded a company about a year and a half ago, to develop small molecule inhibitors of this key enzyme and its substrate because one would predict that such inhibitors would simultaneously activate antitumor immunity and also directly kill the tumor.

DK: I could talk about research all day, but it would be great to switch gears because there is another side to all of this in your life, that I would feel remiss if I didn't mention, and that's that you're considered somewhat of a champion in also women's rights in the STEM-related fields. Could you tell us a little more about initiatives

that you've started, or become involved in recently?

LG: I realized fairly early as my lab grew bigger that we needed to level the playing field for women. Too many women get discouraged and drop out early on because achieving an acceptable work/life balance is challenging to say the least. By the time you're a postdoc, you're thinking about having kids… I mean, I had my three kids when I was 28, 30, and 36. I started my own lab when I was 31. And it was very tough! I wanted to make it easier for my graduate students and postdoctoral fellows. If you provide a technician to a postdoctoral fellow who has young kids, then the work can continue if you have to run out and go to the kids' school, kid's ball game because there is a pair of hands carrying on with the experiments. If your hours are 8 to 6 or 8 to 5, because you have to pick up the kids at daycare or whatever, it's good to have somebody else to continue the experiment. My lab was well-funded, so I was able to provide that for the female grad students and postdocs who had childcare responsibilities. And it worked! They were productive, and a lot of them are in very wonderful positions now. So then I convinced the NIH – the NIAID, in particular, to pilot this program, which they call PCTAS. I was on the Summers's task force for women in science and engineering and was able to suggest some programs that would help support women there as well.

When I arrived at Weill Cornell, I was able to build an excellent childcare center, and made sure we had paid maternity leave and automatic suspension of tenure. It was a pleasure to look for really talented women to develop at Weill Cornell – and there were plenty of them. I found talented junior and senior women faculty whose talents had been unrecognized, and promoted them. It's important to provide intellectual and emotional support, but one also has to provide financial support. I raised a fund for young female faculty who were assistant professors. They had a great idea, they had a great discovery, they needed to push it for all it's worth, they don't have time to wait to get a grant from the NIH. Give them money now so they can hire another postdoc, so that they can push that discovery further, faster. Put them in positions of leadership, because we don't have enough senior leaders who are women who can serve as role models.

DK: So you kind of facilitated that.

LG: Yes, this has been a big priority for me. And not just women – men too. But a special place for women – because I know how hard it was when I went through it. My first husband was a surgeon so I was the primary childcare parent. Three kids, you really learn how to manage your time for sure, but I didn't want the people I trained to have to go through what I went through. At that time I really didn't have a mentor, nobody really helping me – it was tough, it was really hard.

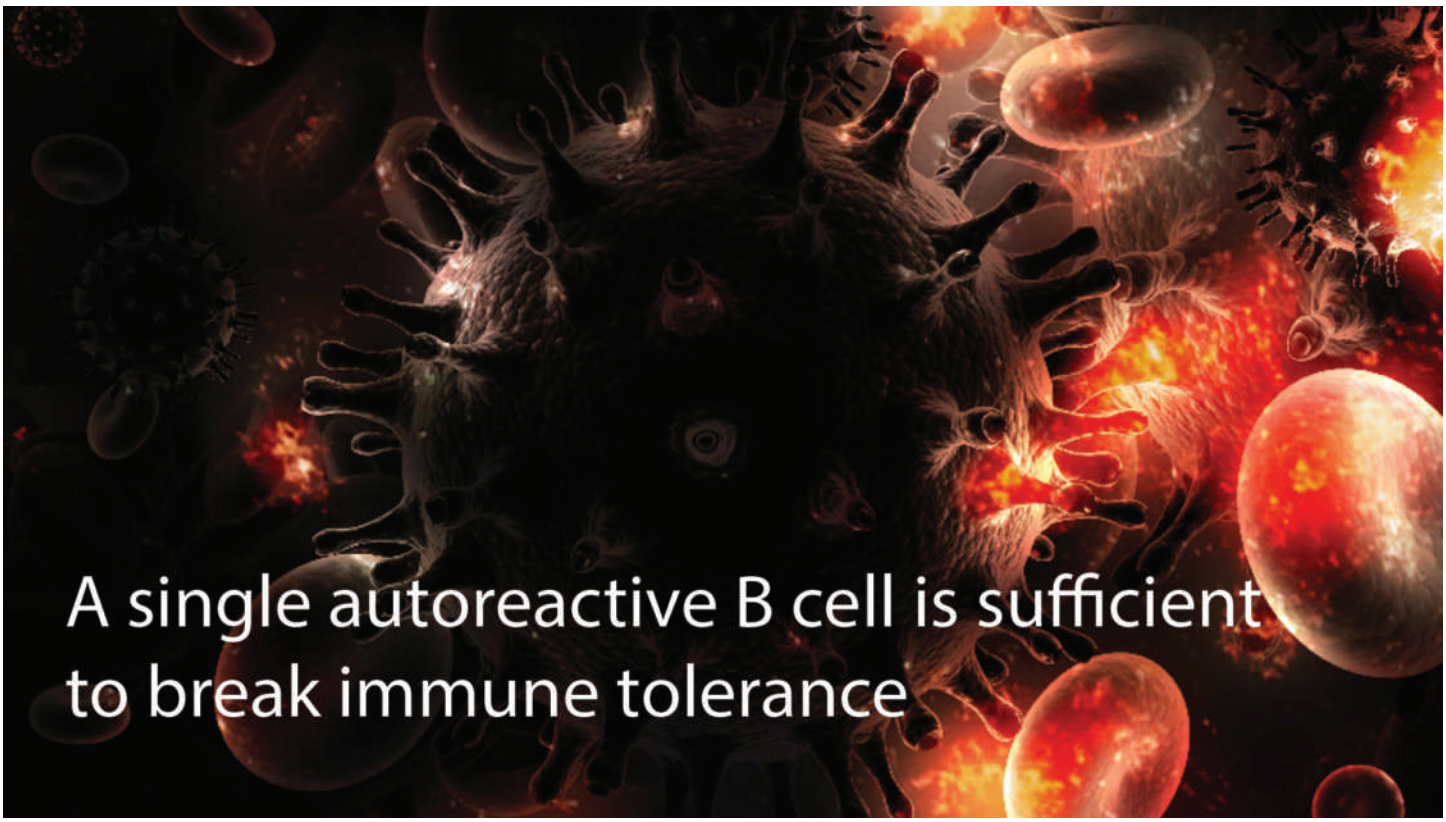
DK: One last question: I was wondering what your thoughts were on how to help before – because a lot of your work has centered on doing great things for women professionals, people who are already established, and I was wondering: do you have any advice for young women who might be struggling with their identity, within the phase of education, within their field of interest, before they become established as a professional? Any dreams or advice you have for female students who might feel gender-based pressure?

One piece of advice is: act as if you're self-confident, even if you aren't.

LG: I think having self-confidence is a very big part of it. And women tend to have less self-confidence, even though it is not deserved. I saw this over and over again in my grad students and postdocs. Certainly in college students as well; I've had a lot of undergraduate students over the years. Women just tend to have less self-confidence. One piece of advice is: act as if you're self-confident, even if you aren't. You have to believe in yourself. And you have to be willing to put yourself out there. Take risks and speak up, for yourself. Men – and this is a gross generalization – but men tend to think they are better than they actually are, and women tend to think that they are not as good as they in fact are. And that's something that has to be worked on, the younger the better.

DK: Thanks for your time.

LG: You're welcome.



A single autoreactive B cell is sufficient to break immune tolerance

Max Miao '19

Systemic lupus erythematosus (SLE) is a chronic autoimmune disease that affects an estimate of 1.5 million people in the US. It currently has no cure and can cause a variety of symptoms at multiple locations, including anemia, joint pain, skin lesion, kidney damage, and neurological disorders. SLE is caused by autoreactive B cells that produce auto-antibodies, antibodies that attack the body's own healthy tissues. Each B cell produces a unique antibody by virtue of their ability to undergo somatic recombination. Thus, B cells that produce different antibodies are called "clones." Many B cell clones exist in the body, and every clone produces a single type of antibody against a single antigen, a molecule that the antibody can detect as foreign. Upon binding to its cognate antigen, an antibody tags it for destruction by other immune cells and triggers inflammation. Although autoreactive B cells are usually eliminated during early developmental stages, they occasionally escape this process and cause a variety of autoimmune diseases, including SLE. The origin of these autoreactive B cells is still largely unknown, and a better understanding of this problem is crucial for finding a cure for SLE.

The diversity of SLE symptoms mirrors the diversity of auto-antibodies. At least 180 types of autoantibodies have been discovered in SLE patients, targeting a wide range of antigens including nucleic acids, cell membrane components, and secreted proteins[1]. These auto-antibodies, however, usually do not arise all at once during disease progression. Anti-DNA autoantibodies, a common molecular marker for SLE, may appear in the bloodstream as early as a few years before any noticeable symptoms[2]. At the onset of disease, B cells producing auto-antibodies against other antigen targets rapidly expand and cause widespread inflammation, a phenomenon called epitope spreading. Understanding the triggers of epitope spreading is crucial for developing more effective therapeutic and preventive measures against SLE.

Recently, a team of scientists led by Michael Carroll at Harvard Medical School offered new insights into the mechanism of epitope spreading in their paper published in Cell[3]. Using a novel mouse model, they discovered that introduction of a single autoreactive B cell clone drives the production of other B cell clones against different auto-antigens. In their study, they used a method called "mixed chimera" to generate a mouse strain that contains both wild-type B cells (WT) and an autoreactive B cell clone (named 564Igi) against a known DNA antigen. To do so, they ablated the immune system of wild-type mice with irradiation, reconstituted their hematopoietic system with marrow from both WT and 564Igi mice, and allowed the immune system to regenerate with the transferred bone marrow. The 564Igi B cells expressed a fluorescent protein, which allowed the authors to track their location and proportion.

Surprisingly, although the chimeric mice showed an increase in total autoreactive antibodies compared to fully wild-type mice, the majority of these antibodies were produced by WT B cells. In support of this finding, the authors observed that 6-8 weeks after bone marrow transfer, almost all B cells were wild-type in germinal centers (GCs), tissue sites where B cells aggregate, mature, and produce antibodies. Analysis of autoantibodies in GCs by sequencing and

antigen binding assay revealed a diverse range of targets, similar to the phenomenon observed in epitope spreading. As a control, the authors also generated a chimeric mouse strain containing WT B cells and a B cell clone expressing an inactive antibody. These mice displayed minimal amounts of auto-antibodies, indicating the chimera experiment itself did not have an effect on autoimmunity. These results suggested that instead of directly contributing to autoimmunity, 564Igi B cells triggered generation of other autoreactive B cells clones.

The authors then investigated whether the 564 Igi B cells were also required for the proliferation and survival of other autoreactive B cells once they were produced. To do so, they generated another batch of chimeric mice similar to the previous ones, except the 564Igi B cells in these mice also expressed a receptor that can bind to tamoxifen. Tamoxifen administration would selectively kill these cells, leaving other cells, including WT B cells, unharmed. Tamoxifen administration effectively eliminated 80-90% of 564Igi B cells, while the total B cell frequency in autoreactive GCs were unaffected. These results suggest that once other autoreactive B cells are induced by one autoreactive strain, the original strain is no required for their proliferation and survival.

Establishment of this link will potentially open doors for more effective treatment against autoimmune diseases.

Despite the diversity of antigens targeted by WT autoreactive B cells, sequencing analysis showed that the antibodies produced by these cells converged on specific sequences even in GCs from different mice. To further study whether the convergence of antibodies is caused by a convergence of B cell clones, the authors constructed another 564Igi-WT chimera mouse model containing 564Igi B cells and WT Confetti B cells. Every Confetti B cell clone expresses one of ten fluorescence markers with different colors, which serves as a marker for their antibody identity. Four weeks following bone marrow transfer, the authors observed a decrease in color diversity in autoreactive GC and dominance of a few B cell clones. In comparison, the Confetti-only control mice without 564Igi B cells showed constant B cell diversity. These results suggest that once autoimmunity is established by 564Igi B cells, WT B cell clones in GCs compete against each other, resulting in the survival of a few clones that are the most autoreactive.

This study conducted by Carroll and colleagues was the first to demonstrate that a specific autoreactive B cell clone can cause expansion of other autoreactive clones. Although epitope spreading had long been observed, it was unclear whether the early autoreactive B cell clone is causally linked to the subsequent expansion of other autoreactive clones. Establishment of this link will potentially open doors for more effective treatment against autoimmune diseases such as SLE. A major hurdle in treating SLE is the heterogeneity of this disease: patients harbor a wide spectrum of autoantibodies against diverse targets, making a one-for-all therapy unlikely[4]. Now that clinicians know that a single autoreactive B cell clone can lead to full-blown disease progression, researchers may develop new strategies to combat SLE by identifying and eliminating this clone at early stages, even before the induction of other autoreactive B cells and presentation of symptoms.

References

[1] Yaniv, Gal, et al. "A volcanic explosion of autoantibodies in systemic lupus erythematosus: a diversity of 180 different antibodies found in SLE patients." Autoimmunity reviews 14.1 (2015): 75-79.

[2] Arbuckle, Melissa R., et al. "Development of autoantibodies before the clinical onset of systemic lupus erythematosus." New England Journal of Medicine 349.16 (2003): 1526-1533.

[3] Degn, Søren E., et al. "Clonal Evolution of Autoreactive Germinal Centers." Cell 170.5 (2017): 913-926.

[4] Eisenberg, Robert. "Why can't we find a new treatment for SLE?." Journal of autoimmunity 32.3 (2009): 223-230.