Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Hallie Averbach Date

3q29 Microdeletion Syndrome: Phenotypic Differences between Urban and Rural Populations

By

Hallie Averbach
Master of Public Health

Environmental Health

W. Michael Caudle, PhD
Committee Chair

Jennifer Gladys Mulle, MHS, PhD
Committee Member

Melissa M. Murphy, PhD
Committee Member

3q29 Microdeletion Syndrome: Phenotypic Differences between Urban and Rural Populations

By

Hallie Averbach

B.S.
University of Michigan
2017

Thesis Committee Chair: W. Michael Caudle, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Environmental Health
2019

Abstract

3q29 Microdeletion Syndrome: Phenotypic Differences between Urban and Rural Populations

By Hallie Averbach

3q29 Deletion Syndrome is a genetic syndrome resulting from a recurrent 1.6 Mb deletion on the long arm of chromosome 3 (Willatt, 2005; Ballif, 2008). The syndrome is rare, with a prevalence of about 1 in 30,000. It is host to several phenotypic outcomes, including autism spectrum disorder, heart defects, and intellectual disability (Baliff, 2008; Glassford et al., 2016). However, the phenotypic outcomes vary among the population, meaning that not everyone with 3q29 Deletion Syndrome is subject to the same phenotypic outcomes. Currently, researchers do not fully understand the reason behind this variability. It may be due to the underlying biological mechanisms associated with the deletion or other factors, such as environmental factors. Research efforts of 3q29 Deletion Syndrome are focused at Emory University through an online registry and an in-person study. The 3q29 Registry has identified 140 cases of 3q29 Deletion. Nineteen of these cases have undergone physical, behavioral, and intellectual assessments, as outlined in Murphy et al. (2018) to learn more about the 3q29 Deletion Syndrome phenotype.

Currently, neither the registry nor the in-person study consider how environmental exposures related to phenotypic outcomes associated with 3q29 Deletion Syndrome. Environmental exposures include exposures to chemicals from soil, air, and water, and other factors, like exposures to stress, nutrition, and access to care. Existing data supports the hypothesis that interaction between genetic factors and environmental exposures influences phenotypic outcomes (Eley et al., 2004; Kim-Cohen et al., 2006; Manuck and McCaffery, 2014). Understanding how the environment interacts with the 3q29 deletion and the associated 3q29 Deletion Syndrome could lead to insights on how environmental factors play a role in how the differences in phenotypic outcomes manifest among the population. By identifying potential environmental exposures related to 3q29 Deletion Syndrome at the broad level of urban versus rural living areas, assessing whether individuals with 3q29 Deletion Syndrome experience different health outcomes based on living area, and suggesting future environmental exposure studies in relation to 3q29 Deletion Syndrome, this analysis aims to develop a preliminary understanding of how environmental factors may play a role in health outcomes associated with 3q29 Deletion Syndrome.

Keywords: 3q29 microdeletion, 3q29 deletion, gene-environment interaction, urban and rural populations, environmental exposures, access to care

3q29 Microdeletion Syndrome: Phenotypic Differences between Urban and Rural Populations

By

Hallie Averbach

B.S.
University of Michigan
2017

Thesis Committee Chair: W. Michael Caudle, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Environmental Health
2019

Table of Contents

Introduction

3q29

Background and prevalence

3q29 Deletion Syndrome is a rare genetic disorder with a prevalence of 1 in 30,000 individuals. It results from a recurrent 1.6 Mb deletion on the long arm of chromosome 3 (Mulle, 2010). It is unclear as to why or how the deletion occurs (Mulle, 2010). First discovered in 2005, this deletion results in deletion of 21 genes in the region, as summarized in Table 1. Not all the functions of genes in the region are known, but some, such as FBX045, are associated with synapse formation, axon pathfinding, and neuronal migration. Others, such as PAK2 are involved in mediating molecular processes such as neuron migration (Mulle et al., 2017). The deletion of these genes could be related to the phenotypic outcomes associated with the syndrome. This syndrome is a result of the deletion, which often appears de novo. It is associated with a range of physical, psychosocial, and neurological outcomes. Some of these outcomes include heart defects, dental abnormalities, autism, anxiety, and intellectual disability (Baliff, 2008; Glassford et al., 2016).

Phenotype

The 3q29 deletion is characterized by clinical features of developmental delay or intellectual disability (Mulle et al., 2016). Individuals with the syndrome present a higher rate of autism spectrum disorder and anxiety diagnoses than the general population (Glassford et al., 2016). Most notably, individuals with 3q29 Deletion Syndrome have a 40 fold risk increase for schizophrenia, which makes it the largest molecular risk for schizophrenia (Mulle, 2015; Stefansson et al., 2014). Additionally, individuals with 3q29 Deletion

Syndrome tend to exhibit these neurophysiological behaviors earlier than the average age of risk (Glassford et al., 2016). The syndrome is also associated with a variety of physical health outcomes such as heart defects, dental abnormalities, and low birthweight (Glassford et al., 2016).

Diagnosis, surveillance, and treatment

The 3q29 deletion is diagnosed by detection of the 1.6 Mb heterozygous deletion at the approximate position of chr3:195998129-197623129 in the reference genome (Mulle et al., 2016). This detection is typically done via chromosomal microarray. After diagnosis, surveillance is necessary for those affected by 3q29 Deletion Syndrome. Areas of surveillance include: feeding and nutrition patterns, progress in meeting developmental milestones, cognitive development, and possible neuropsychiatric manifestations (Mulle et al., 2016).

Due to the health outcomes associated with 3q29 Deletion Syndrome, treatment is often needed to manage these outcomes. Because of global developmental delay, those impacted by 3q29 Deletion Syndrome may be subjected to early speech and language therapy to address speech delays or physical therapy address fine and gross motor delays (Mulle et al., 2016). An individualized education program (IEP) may be developed for school-age children.

To treat neuropsychiatric disorders, care by a child psychiatrist/psychologist may be necessary. To address problems with feeding and nutrition patterns, families may consider feeding therapy and consideration of gastrostomy tube. Physician visits may be required for routine management of congenital heart defects. Dental work may be needed for dental abnormalities.

Past and current efforts to understand 3q29 Deletion Syndrome

3q29 Deletion Syndrome was first described in 2005 by Willatt et al. (2005). In this study, the researchers characterized six patients with almost identical deletion sizes. Since then, efforts have been made to learn more and better understand the syndrome. A registry, described in Glassford et al., (2016) exists for individuals with 3q29 associated syndromes, including 3q29 Deletion Syndrome. Additionally, a study at Emory University, described in Murphy et al. (2018), is working on characterizing the 3q29 Deletion Syndrome phenotype. Currently, this study is the only one of its kind.

As previously noted, despite the genetic shared variant of the 3q29 deletion, variability exists between phenotypes and in the severity to which the different phenotypic traits are expressed. The reasons for variability are unknown and could be due to many factors, including environmental factors. For example, we know that environmental chemicals exposure is a risk factor for autism spectrum disorder (Becker, 2007). Autism spectrum disorder is one example of a phenotype of 3q29 Deletion Syndrome that could be modified by other factors besides the gene itself. Looking at other factors, including environmental exposures such as air quality, water quality, access to care, and presence of chemicals in the environment, could lead to understanding more about how these phenotypes arise or the variability at which they come about.

Gene-environment interaction

Gene-environment interaction can explain how individuals with the same genotype are subjected to different phenotype (Purcell, 2002). Evidence indicates that the environment

impacts and modifies how genes are presented as a phenotype (Eley et al., 2004; Kim-Cohen et al., 2006; Manuck and McCaffery, 2014).

Mental health

Gene-environment interactions have been analyzed in many studies related to mental health. Studies have indicated how the environment can affect serotonin receptors and make individuals more susceptible to depressive symptoms and diagnoses based on where they are living (Jokela, Lehtimäki, and Keltikangas-Järvinen, 2007). In this study, individuals with the T allele of the T102C polymorphism of the HTR2A gene, responsible for the receptor 2A, were more likely to experience depressive symptoms in rural areas than those with the T allele living in urban areas. However, there was no difference with the C allele of this gene. In this example, it was speculated that the differences in expression of the T allele were potentially due to the fact those living in urban settings had a higher education level and a larger social network for support. This analysis sheds light on some of the environmental factors that can lead to outcomes related to mental health in those that are genetically predisposed to that outcome.

This concept is present in other studies, as well. Eley et al., (2004) also looked at gene-environment interactions in the case of developing depression. Depression is a complex condition that is influenced by many genetic and environmental factors. The researchers point out that in the case of many neuropsychological disorders, a combination of genetic risk factors and environmental risk factors work together to develop the phenotype and determine the severity of the disorder, making the whole situation complex. This study specifically looked at the effect that exposures to stress as well as the 5HT serotonin receptor gene. This study found that the 5HT gene was more directly related to depression outcomes,

and environmental exposures were not significantly associated with depression outcomes. Nonetheless, this analysis examines how complex neuropsychological disorders are in terms of genetic and environmental factors and how these interactions can complicate the expected outcomes.

Similar evidence has been exhibited for other mental health conditions, such as anxiety. This was determined in a study from Kender et al. (1986), where researchers considered fourteen symptoms related to anxiety and depression among 3,798 sets of volunteer twins in Australia. This study found that for most of the symptoms, the outcome was dependent on genetic factors and influenced by environmental events specific to the individual, such as undue stress from the death of a parent. For four of the symptoms, the researchers were unable to distinguish between whether genetic or environmental factors played a larger role in the phenotypic outcome. The study concluded that genetic and environmental factors are sometimes so intertwined that it is difficult to isolate just one to determine the etiology of the condition. This shows how complex these gene-environment interactions can be, especially when the outcome is related to mental health.

These complexities are not unlike the complexities associated with 3q29 Deletion Syndrome. Additionally, considering that the syndrome is pleiotropic, different environmental factors may have different outcomes on different phenotypes associated with the condition. Looking at these concepts through the lens of 3q29 Deletion Syndrome can help understand how the environment interacts with this mutation to present some of these phenotypic outcomes.

Urban and rural settings

Whether someone lives in an urban or rural area changes the factors they are exposed to, including both environmental and social factors. Some environmental factors include exposure to different chemicals, different levels of air quality, and access to clean water. Some social factors include socioeconomic status and access to healthcare. This is by no means an exhaustive list, but just a sample to show how many differences exist between urban and rural settings. All of these factors play a role in the health outcomes observed in the individual.

Chemical Exposures

One difference between urban and rural settings is exposure to chemicals, such as pesticides. Those living in rural settings may be exposed to higher levels of pesticides, which are known to have effects on neurological development. Furthermore, a study that looked at the Texas Birth Defects Registry found that babies born to mothers living in rural settings have a higher incidence of being born with a heart defect (Langlois, et al., 2009). The researchers speculate that the higher prevalence of agricultural chemicals present in rural communities impacts on the fetus's heart development. However, chemical exposures also exist in urban settings and can contribute to developmental outcomes. A study from China found that dietary cadmium exposure was higher in those living in city settings than in rural settings (Watanabe et al., 1998). Additionally, a study in Toronto found that urban settings had levels of persistent organic pollutants 5-10 times higher than a rural area 75 kilometers north of the city (Harner et al., 2004).

Air Quality

Additionally, living in an urban setting could expose someone to lower levels of air quality. This could be due to increased pollution and particulate matter in the air from increased traffic patterns. A study in Switzerland found that levels of particulate matter in rural areas were consistently a third lower than levels in urban areas (Monn et al., 1995). Another study shows that increased population size and increased rates of urbanization leads to increased levels of air pollution, specifically increased carbon dioxide emissions (Cole and Neumayer, E. 2004).

Access to Care

Due to these different environmental exposures, health outcomes among urban and rural populations may differ. Beyond environmental exposures, the type of care one has access to varies based on the setting (Hartley, 2004). More than 61 million people living in rural settings in the United States do not have adequate access to care (Eberhart et al., 2001). Without this care, both diagnosis and treatment is sparse. For example, a study from Australia found that congenital heart defects were more common closer to urban centers (Kidd, Lancaster, and McCredie, 1993). The researchers believe that this is due to the fact that those living in more urban areas have increased access to care and would be more easily diagnosed.

Phenotypic outcomes

These differences translate to difference in phenotypic outcomes. For example, exposure to different factors in urban and rural areas has an effect on the birthweight of newborns (Reading, Raymond, and Jarvis, 1993). Additionally, studies (Kelleher, 1993; Hannaway, 1970) found that low birth weight was associated with failure to thrive. Since

living in an urban area is associated with increased risk of low birth weight, these individuals may also be at an increased risk for failure to thrive. Beyond this, rates of dental abnormalities differ between urban and rural settings (Sanadhya et al., 2015; Levin et al., 2010; Olasoji and Odusanya, 2000). Observed differences between rural and urban settings is also true for mental health outcomes, such as anxiety (Vassos et al., 2016; Vega et al., 1998).

However, the link between urban and rural settings and different phenotypic outcomes are not always clear cut. For example, some studies indicate that there is an urban-rural distribution of autism and that autism spectrum disorder diagnosis rates are higher in urban settings, which could be influenced by differences in access to care (Gillberg, 1984; Williams, Higgins, and Brayne, 2006; Becker, 2007). In fact, due to heightened mercury exposure and other environmental toxins in urban settings, increased urbanization is considered to be a risk factor for autism (Becker, 2007). However, exposure to insecticides in rural communities have also been found to be linked to the development of autism (Shelton, Hertz-Picciotto, and Pessah, 2012; Landrigan, Lambertini, and Birnbaum, 2012; Pessah and Lein, 2008). There is speculation that exposure to neurotoxins, such as insecticides and other industrial chemicals, during the period of brain development can lead to neurodevelopmental disorders, including autism, in the offspring (Grandjean and Landrigan, 2006). Although the exact exposure of chemicals differs between urban and rural environments, cases have been made among both settings to show that exposure to neurotoxins have led to an increased incidence of autism in both areas.

It is clear that differences between urban and rural settings are widespread and complex. Due to this, the present study will analyze of how environmental factors play a role in 3q29 Deletion Syndrome will look at urban and rural differences broadly. This project will

help to fill the knowledge gap between the health outcomes associated with 3q29 Deletion Syndrome and how they manifest themselves in urban and rural settings. Future analyses could look at some of these factors more in depth.

The hypothesis being tested is that there are differences in phenotype between those who have 3q29 Deletion Syndrome and live in a rural setting compared to those who have 3q29 and live in an urban setting. The prediction is that there will be a higher prevalence of autism spectrum disorder, dental anomalies, anxiety, and depression among those who live in urban areas. Birth weight is also expected to be lower in urban areas. History of seizures and reports of failure to thrive are expected to be higher in rural locations.

Methods

Registry

The 3q29 Registry (3q29deletion.org) provides extensive self-reported information from study subjects and their parents or guardians on individuals with 3q29 syndromes including outcome data regarding physical, behavioral, and psychological characteristics (Glassford et al., 2016). Families of individuals with 3q29 syndromes or the individuals themselves usually find the registry from clinical genetics reports or by performing internet searches for “3q29” and clicking on the registry link. The registry has been active since 2013 and is still accepting enrollment.

Study sample

All registry participants who registered by October 4, 2018 were included in the initial download (n=278). Twenty-two individuals who did not report a zip code were excluded. As the registry consists of 3q29 deletion carriers, twenty-seven 3q29 duplication

carriers, and eighty-two control participants, only those who have the 3q29 deletion were included. Since all the information pertinent to the study is in the medical questionnaire, thirty individuals who did not complete that questionnaire were excluded.

Finally, five instances occurred where the same participant had multiple entries in the registry. This was mostly due to both parents entering data for the participant, or someone forgetting their registry password and beginning a new entry later. Duplicate entries for the same individual were compared using a match feature in Microsoft Excel. Any variables that differed between entries were counted as missing. The final sample size of 107 individuals included participants from both the United States and international participants. There are 63 males and 44 females in the sample. The mean age of individuals in the sample is 13 years old. Data are stored as a Microsoft Excel file.

Variables

The variables in the download contain a subset of the medical questionnaire as well as demographic information. The variables downloaded included data pertaining to both psychological (anxiety diagnosis and autism spectrum disorder diagnosis,) and physical (birth weight, failure to thrive, whether the individual was born with a heart defect, presence of dental abnormalities) phenotypes. In the registry, birth weight is reported categorically in one-pound increments. In this study, the midpoint of each increment was used for the analysis. Since the mean age of individuals in the dataset was 13 years old, physical developmental characteristics and ASD are apparent. However, this is still too young for some other psychological phenotypes associated with 3q29, such as depression or schizophrenia, to be diagnosed. Therefore, these variables were not included in the final analysis. Information of potential confounding variables, such as gestational age, sex,

ethnicity, and race, were also captured in the dataset. Calculated variables included population density of current city, urban or rural classifications, distance from city of birth from current city, and distance to hospital from current city.

Classification

These data were categorized into urban or rural classifications by using participants' zip codes. The U.S. Census defines an urban area as one that has a population density of over 1,000 people per square mile (Ratcliffe et al., 2016). For participants in the United States, SimplyAnalytics was used to determine population density for each participants' reported zip code (SimplyAnalytics, 2018). For international participants, zip codes and cities were used to look up population density data for that area based on information provided by the country. The population density of all participants was compared to the U.S. Census definition of urban to determine if a participant lived in a rural or urban area, regardless of if they lived in the United States or outside of the United States. Both population density and the urban and rural classifications were included in the final dataset.

To determine how far participants live from where they were born, Google Maps was used to provide directions from the current zip code to the recorded city of birth. These data were stored in miles. This information was used to determine if exposure data could be generalized across the lifespan of the participants.

To see how far individuals lived from a hospital, zip codes were entered into Google Maps. Directions were calculated from this zip code to a generic search query of "hospital," which would automatically pull up the closest hospital to the individual. This distance, in miles, was recorded in the dataset. This variable was used as a proxy to estimate access to

care among individuals and see if there was some difference in access to care between those who live in urban or rural settings, as that may be a confounding variable.

Analysis plan

All data analysis was performed using SAS 9.4 software (SAS Institute Inc, 2013). Descriptive statistics including mean and standard deviation stratified by urban and rural setting were reported for age, birthweight, and distance to nearest hospital. Unpooled t-tests were used to determine if there was a true difference in distance to the closest hospital between urban and rural populations. Because the variances of distance to hospital were not equal between urban and rural settings, unpooled t-tests were used to analyze the significance of this potential difference. Prevalence of anxiety, autism, failure to thrive, heart defects, and dental abnormalities was also reported and stratified by urban and rural setting.

Regression models were used to analyze the relationship between phenotypic outcomes of 3q29 Deletion Syndromes and how the urbanization of a person's environment. A linear regression model was used to determine the relationship between birthweight and urbanization, controlling for sex and gestational age. Logistic regression models were used to determine the relationship between urbanization and failure to thrive rates, anxiety diagnoses, autism diagnoses, dental abnormalities, and heart defects. Since autism rates are higher among males (Loomes, Hull, and Mandy, 2017) and diagnosis is contingent upon age, this model controlled for age and sex. Anxiety diagnoses are also contingent upon age (Kessler et al., 2007) so the anxiety model also controlled for age. The variables requiring a diagnosis (failure to thrive, heart defect, dental abnormalities, anxiety, and autism) were also controlled for distance to medical center, since those living closer to a medical center may have more

access to care and may be more likely to receive a diagnosis and treatment. An alpha of 0.05 was set prior to running any analyses.

Results

This analysis sought to examine the phenotypic differences between individuals with 3q29 Deletion Syndrome living in urban areas and individuals with 3q29 Deletion Syndrome living in rural areas. In order to do this, a study sample from the 3q29 Registry was extracted and analyzed. Participants were grouped into urban and rural populations based on U.S. Census parameters of urban areas and the participants' zip codes.

As shown in Table 2, the study sample consisted of 107 individuals with 3q29 Deletion Syndrome. Of these individuals, 63 were male and 44 were female. The average age of an individual in this sample was 13.

Thirty-one registry entries included a self-report of autism. Thirty registry entries included a self-report of anxiety. Twenty-seven reported heart defects. Ninety-one reported dental abnormalities. Thirty-seven reported failure to thrive. The mean birth weight in the population was 6.22 pounds.

56 of these individuals lived in urban settings, as defined in the methods section, and 51 individuals lived in rural settings. Table 2 shows the average age of individuals with 3q29 living in an urban population was 9.93 years old (standard deviation=6.00) and the average age of individuals with 3q29 living in a rural population was 15.87 years old (standard deviation=13.458).

As shown in Table 3, the percent of individuals with anxiety diagnoses in urban areas (24.1%) was 10.6% lower than the percent of individuals with anxiety diagnoses in rural areas (34.7%). The percent of individuals who presented dental abnormalities was 9.8%

higher in rural (90.2%) than urban (80.4%). However, failure to thrive diagnoses were 10.3% higher in urban areas (40.7%) than rural areas (30.4%). The prevalence of autism was 0.4% higher in rural areas (30%) than urban areas (29.6). Heart defects were 3.3% higher in urban (26.8%) than rural (23.5%) populations (Figure 1). Table 4 shows the mean birth weight was 0.1 pound lower for urban populations (6.2 lbs.) as compared to rural (6.3 lbs.) populations (Figure 2).

The average distance to a hospital for those living in an urban setting was 5.98 miles while the average distance to a hospital for those living in a rural setting was 18.3 miles (Figure 3). After conducting an unpooled Satterhwaite t-test, this difference in distances was found to be significant ($p=0.0086$). This shows that the distance someone with 3q29 living in a rural setting must travel to receive medical care is significantly farther than the distance someone with 3q29 living in an urban setting has to travel to receive the same care. Therefore, distance to a hospital could be a potential covariate in the differences in phenotypes among those with 3q29 Deletion Syndrome.

Table 5 shows results of regressions with living in an urban environment as the reference point. Based on this analysis, living in an urban or rural setting did not have a significant effect upon the differences in phenotypic outcomes associated with 3q29 Deletion Syndrome. Additionally, after controlling for distance to nearest hospital as a proxy for controlling for access to care, no significant difference was found between those living in urban or rural settings and phenotypic outcomes associated with 3q29 Deletion Syndrome (Table 6).

Discussion

This study was designed to test the hypothesis that there is some difference in phenotype between those who have 3q29 and live in a rural setting compared to those who have 3q29 and live in an urban setting. In this analysis, individuals with 3q29 Deletion Syndrome from the 3q29 Registry were categorized into living in an urban or rural setting by the population density of their zip codes. Regression analyses were used to determine whether there was an association between the type of setting in which a person lives and six phenotypic characteristics of 3q29 Deletion Syndrome (autism, anxiety, failure to thrive, heart defect, dental abnormality, and low birth weight). Although no regression models determined statistical significance between the relationship of the setting in which someone with 3q29 Deletion Syndrome lives and their phenotypic outcomes, it was found that individuals living in a rural setting lived, on average, 18.3 miles farther from a hospital than those living in an urban setting. This analysis helps begin to understand how environmental factors, including environmental exposures and access to care, may play a role in development of the 3q29 Deletion Syndrome phenotype.

This analysis used population density as a proxy for urbanization. However, using population density as a proxy and separating individuals into urban and rural areas is not specific enough of a classification and may cause some of the effects to get lost. Instead, it may be beneficial to look at the aspects of the environment that differ between these two settings on their own, such as exposure to chemicals, air quality, or water quality, as opposed to the broad classification. For example, as previously mentioned with hypothesis around the development of autism, neurotoxins that are thought to contribute to a higher risk of autism are found in both urban and rural environments (Becker, 2007; Grandjean and Landrigan, 2006). Furthermore, other environmental factors, such as nutrition of the mother, or stress

and trauma, cannot accurately be predicted by classifying an area as urban or rural, but influence the health outcome of the offspring. Studies have shown that maternal stress alters the epigenome and can have negative health effects on the offspring (Tyrka et al., 2012). This effect, along with many others, would not be picked up in a population density analysis, but is still an environmental exposure that alters the health outcome of an individual. Therefore, looking at the effects of these other factors more closely may shed better insight on differences in health outcomes based on environmental exposures.

Furthermore, transprogeneral theory could play a role in phenotype of 3q29 Deletion Syndrome. As noted in earlier studies (Skinner, 2011; Perera and Herbstman, 2011; Heijmans et al., 2008), exposures during the parent's lifespan can create epigenetic differences that are passed onto the offspring. Therefore, looking at the parents' past exposures could be helpful. Potentially, the mother or father were exposed to some environmental factor and changes were made to their gametes. Later, the combination of these changes and occurrence of 3q29 Deletion Syndrome could lead to variability in phenotype. Overall, environmental factors that modify gene expression, both in the individual with 3q29 Deletion Syndrome and in that person's parents, could be an explanation to variety in phenotype and this idea worth pursuing more.

Beyond the importance of further scrutinizing the differences between urban and rural factors, this analysis shed light on the differences in access to care between urban and rural populations of individuals with 3q29 Deletion Syndrome. On average, individuals with 3q29 Deletion Syndrome living in rural areas had to drive 12.3 miles further to get to a hospital than individuals with 3q29 Deletion Syndrome living in urban areas (rural=18.3 miles, urban=5.98 miles). However, what is interesting, is that despite this difference, there do not

appear to be any significant differences in prevalence of aspects of the 3q29 Deletion Syndrome phenotype (autism, anxiety, dental abnormalities, heart defect, failure to thrive, or low birth weight) between people living in urban and rural areas. That means that even with people in urban areas being way closer to hospitals than those in rural areas, there are about equal numbers of urban and rural populations being diagnosed with the same conditions. We would likely expect that individuals living closer to a hospital would be more likely to be diagnosed with conditions compared to those living further away, but that does not seem to be the case in this population.

Perhaps those living in rural areas and further away from hospitals still have good access to care, meaning that they are able to get to a healthcare provider in and receive the care they need in a timely manner. Conversely, it is also possible that because the phenotypes exhibited in this population are more severe, therefore are more likely to be identified and require more medical attention than the average population, so parents or caregivers are more likely to seek out care, even if that means traveling further to see a healthcare provider.

Limitations

This study has limitations associated with it. First, all of the data is from a survey and it is self-reported. Self-reported data is subject to inaccuracies as participants may try to answer in such a way as to portray themselves in a certain manner. Additionally, reports of historical data is subject to recall bias, which could contribute to inaccuracies in the data.

Additionally, since there is no follow up data, this study consisted of a cross sectional analysis. Due to the nature of cross-sectional studies, there is no way to determine causality of exposures to outcomes. This means that even if there were some associations between urban and rural settings compared to phenotypic outcomes, there is no way of determining if

those outcomes were due to someone living in an urban or rural setting or if they were due to another factor.

Another limitation to this study is that the small sample size leads itself to underpowered results. Because the sample size was only 107 participants and they were divided into two groups, there was not enough power to determine if a meaningful association actually existed within the population. Increasing the sample size would increase the power of the study without decreasing the alpha.

Finally, this analysis uses population density to classify and define urban or rural settings. However, since the differences between urban and rural settings that can change gene expression are often related to differences in exposure, population density may not be the best way to classify settings. Additionally, categorizing individuals in this way contributed to decreasing the power of the study. Nonetheless, this was the most reasonable proxy to use with the data available. Future studies should look to collect more data appropriate to exposure assessment to avoid continuing to use population density as a proxy.

Future Directions

This analysis gives a starting point to future analyses of 3q29 Deletion Syndrome and environmental exposures. Different analysis techniques, such as a biomarker analysis, would allow researchers to study how environmental factors determine phenotypic outcomes of 3q29 Deletion Syndrome more in depth. Additionally, researchers could target analyses more to analyze specific of environmental exposures. Administering a healthcare access and utilization survey would also be useful to better understand non-genetic factors that may affect phenotypic outcome of 3q29 Deletion Syndrome.

Studies on 3q29 Deletion Syndrome could benefit greatly from biomarker analysis in assessing chemical exposures. Biomarkers, such as blood, urine, hair, or fingernails, are widely used and relatively easy to obtain, with blood being one of the most difficult. The Emory 3q29 Project already collects blood as part of its protocol (Murphy et al., 2018), so some serum collection could be set aside for biomarker analysis. Looking at blood samples allows researchers to see different metabolites in the blood, giving light to potential exposures. Hair and nail samples could also be used in biomarker analysis to see chemical exposure over time. Collecting these samples are likely even easier than collecting a blood sample, and still give some sort of longitudinal exposure data.

Additionally, those studying 3q29 Deletion Syndrome could benefit from studying differences in environmental exposures beyond just urban and rural data. As previously mentioned, there are many environmental factors that differ between urban and rural environments, such as air quality, water quality, and chemical exposures. Furthermore, environmental exposures, such as stress, trauma, and nutritional factors, are not dependent upon urban or rural settings, but still may play a role in gene expression and should be considered.

In order to look at these factors more closely, researchers can take a few different steps. First, looking at biomarker analyses, as previously suggested, would give a good idea of what chemical exposures someone has been subjected to. Additionally, having a better idea of location data over time, such as knowing where the individual with 3q29 Deletion Syndrome lived throughout their life or where their mother lived during pregnancy, would give more data points to do a more thorough data analysis of air or water quality data someone was subjected to throughout their life, including in utero. A driving factor as to why

the current analysis only used urban and rural data was because little location data has been reported to date, so any other type of analysis was not very feasible.

Finally, administering a healthcare access and utilization survey would be useful in determining differences in access to care among the population of individuals with 3q29 Deletion Syndrome. These types of surveys have been useful in other studies to show how access to healthcare differs among populations studied (Goodridge et al., 2010; Haggerty et al., 2007; Casey et al., 2001). Administering a survey of this nature could give more insight into how individuals with 3q29 Deletion Syndrome living in rural areas are able to have similar diagnostic rates of associated conditions as those living in rural areas, despite living much farther away from a hospital.

Conclusions

In this hypothesis generating study, a subset of the 3q29 Registry was analyzed to determine how phenotypic outcomes of 3q29 Deletion Syndrome differed between urban and rural populations. Although none of the variables examined (anxiety, autism, birth weight, heart defects, failure to thrive, or dental abnormalities) were statistically different among urban and rural populations, this analysis gives direction for future studies involving environmental exposures and 3q29 Deletion Syndrome. Future studies should consider looking at more specific environmental exposures, such as metal exposures, air quality, or maternal stress. Additionally, since there was a statistically significant difference among distance from a hospital between urban and rural populations, but incidence of health outcomes did not differ between populations, looking further at healthcare access and utilization may be of interest. This analysis is not conclusive or meant to imply causality but

will help prioritize future research in areas pertaining to environmental exposure, including access to care, and phenotypic outcomes of 3q29 Deletion Syndrome.

Bibliography

Adami, P., Duncan, T. M., McIntyre, J. O., Carter, C. E., Fu, C., Melin, M., ... & Fleischer, S. (1993). Monoclonal antibodies for structure-function studies of (R)-3-hydroxybutyrate dehydrogenase, a lipid-dependent membrane-bound enzyme. *Biochemical Journal*, *292*(3), 863-872.

Ballif, B. C., Theisen, A., Coppinger, J., Gowans, G. C., Hersh, J. H., Madan-Khetarpal, S., ... & McDonald, M. (2008). Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. *Molecular Cytogenetics*, *1*(1), 8.

Becker, K. G. (2007). Autism, asthma, inflammation, and the hygiene hypothesis. *Medical hypotheses*, *69*(4), 731-740.

Casey, M. M., Call, K. T., & Klingner, J. M. (2001). Are rural residents less likely to obtain recommended preventive healthcare services?. *American journal of preventive medicine*, *21*(3), 182-188.

Cole, M. A., & Neumayer, E. (2004). Examining the impact of demographic factors on air pollution. *Population and Environment*, *26*(1), 5-21.

Ehrenberg, H. M., Dierker, L., Milluzzi, C., & Mercer, B. M. (2003). Low maternal weight, failure to thrive in pregnancy, and adverse pregnancy outcomes. *American journal of obstetrics and gynecology*, *189*(6), 1726-1730.

Eley, T. C., Sugden, K., Corsico, A., Gregory, A. M., Sham, P., McGuffin, P., ... & Craig, I. W. (2004). Gene–environment interaction analysis of serotonin system markers with adolescent depression. *Molecular psychiatry*, *9*(10), 908.

Gillberg, C. (1984). Infantile autism and other childhood psychoses in a swedish urban region. Epidemiologiccal aspects. *Journal of Child Psychology and Psychiatry*, *25*(1), 35-43.

Glassford, M. R., Rosenfeld, J. A., Freedman, A. A., Zwick, M. E., Unique Rare Chromosome Disorder Support Group, & Mulle, J. G. (2016). Novel features of 3q29 deletion syndrome: results from the 3q29 registry. *American Journal of Medical Genetics Part A*, *170*(4), 999-1006.

Goodridge, D., Lawson, J., Rennie, D., & Marciniuk, D. (2010). Rural/urban differences in health care utilization and place of death for persons with respiratory illness in the last year of life. *Rural Remote Health*, *10*(2), 1349.

Grandjean, P., & Landrigan, P. J. (2006). Developmental neurotoxicity of industrial chemicals. *The Lancet*, *368*(9553), 2167-2178.

Haggerty, J. L., Roberge, D., Pineault, R., Larouche, D., & Touati, N. (2007). Features of primary healthcare clinics associated with patients' utilization of emergency rooms: urban–rural differences. *Healthcare Policy*, *3*(2), 72.

Harner, T., Shoeib, M., Diamond, M., Stern, G., & Rosenberg, B. (2004). Using passive air samplers to assess urban− rural trends for persistent organic pollutants. 1. Polychlorinated biphenyls and organochlorine pesticides. *Environmental science & technology*, *38*(17), 4474-4483.

Hartley, D. (2004). Rural health disparities, population health, and rural culture. *American Journal of Public Health*, *94*(10), 1675-1678

Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., ... & Lumey, L. H. (2008). Persistent epigenetic differences associated with prenatal

exposure to famine in humans. *Proceedings of the National Academy of Sciences*, *105*(44), 17046-17049.

Jokela, M., Lehtimäki, T., & Keltikangas-Järvinen, L. (2007). The influence of urban/rural residency on depressive symptoms is moderated by the serotonin receptor 2A gene. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *144*(7), 918-922.

Kelleher, K. J., Casey, P. H., Bradley, R. H., Pope, S. K., Whiteside, L., Barrett, K. W., ... & Kirby, R. S. (1993). Risk factors and outcomes for failure to thrive in low birth weight preterm infants. *Pediatrics*, *91*(5), 941-948.

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustun, T. B. (2007). Age of onset of mental disorders: a review of recent literature. *Current opinion in psychiatry*, *20*(4), 359.

Kidd, S. A., Lancaster, P. A. L., & McCredie, R. M. (1993). The incidence of congenital heart defects in the first year of life. *Journal of paediatrics and child health*, *29*(5), 344-349.

Kim-Cohen, J., Caspi, A., Taylor, A., Williams, B., Newcombe, R., Craig, I. W., & Moffitt, T. E. (2006). MAOA, maltreatment, and gene–environment interaction predicting children's mental health: new evidence and a meta-analysis. *Molecular psychiatry*, *11*(10), 903.

Landrigan, P. J., Lambertini, L., & Birnbaum, L. S. (2012). A research strategy to discover the environmental causes of autism and neurodevelopmental disabilities.

Levin, K. A., Davies, C. A., Douglas, G. V., & Pitts, N. B. (2010). Urban-rural differences in dental caries of 5-year old children in Scotland. *Social science & medicine*, *71*(11), 2020-2027.

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(6), 466-474.

Manuck, S. B., & McCaffery, J. M. (2014). Gene-environment interaction. *Annual review of psychology*, *65*, 41-70.

Monn, C., Braendli, O., Schaeppi, G., Schindler, C., Ackermann-Liebrich, U., Leuenberger, P. H., & Sapaldia Team. (1995). Particulate matter< 10 μm (PM10) and total suspended particulates (TSP) in urban, rural and alpine air in Switzerland. *Atmospheric Environment*, *29*(19), 2565-2573.

Mulle, J. G. (2015). The 3q29 deletion confers> 40-fold increase in risk for schizophrenia. *Molecular psychiatry*, *20*(9), 1028.

Murphy, M. M., Burrell, T. L., Cubells, J. F., España, R. A., Gambello, M. J., Goines, K. C., ... & Russo, R. L. S. (2018). Study protocol for The Emory 3q29 Project: evaluation of neurodevelopmental, psychiatric, and medical symptoms in 3q29 deletion syndrome. *BMC psychiatry*, *18*(1), 183.

Olasoji, H. O., & Odusanya, S. A. (2000). Comparative study of third molar impaction in rural and urban areas of southwestern nigeria. *TROPICAL DENTAL JOURNAL*, 25-28.

Perera, F., & Herbstman, J. (2011). Prenatal environmental exposures, epigenetics, and disease. *Reproductive toxicology*, *31*(3), 363-373.

Pessah, I. N., & Lein, P. J. (2008). Evidence for environmental susceptibility in autism. In *Autism* (pp. 409-428). Humana Press.

Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research and Human Genetics*, *5*(6), 554-571.

Reading, R., Raybould, S., & Jarvis, S. (1993). Deprivation, low birth weight, and children's height: a comparison between rural and urban areas. *BMJ*, *307*(6917), 1458-1462.

Sanadhya, S., Aapaliya, P., Jain, S., Sharma, N., Choudhary, G., & Dobaria, N. (2015). Assessment and comparison of clinical dental status and its impact on oral health-related quality of life among rural and urban adults of Udaipur, India: A cross-sectional study. *Journal of basic and clinical pharmacy*, *6*(2), 50.

Shelton, J. F., Hertz-Picciotto, I., & Pessah, I. N. (2012). Tipping the balance of autism risk: potential mechanisms linking pesticides and autism. *Environmental health perspectives*, *120*(7), 944-951.

Skinner, M. K. (2011). Environmental epigenetic transgenerational inheritance and somatic epigenetic mitotic stability. *Epigenetics*, *6*(7), 838.

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., ... & Tost, H. (2014). CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*, *505*(7483), 361.

Tyrka, A. R., Price, L. H., Marsit, C., Walters, O. C., & Carpenter, L. L. (2012). Childhood adversity and epigenetic modulation of the leukocyte glucocorticoid receptor: preliminary findings in healthy adults. *PloS one*, *7*(1), e30148.

Watanabe, T., Zhang, Z. W., Qu, J. B., Xu, G. F., Song, L. H., Wang, J. J., ... & Ikeda, M. (1998). Urban–rural comparison on cadmium exposure among general populations in Shandong Province, China. *Science of the total environment*, *217*(1-2), 1-8.

Willatt, L., Cox, J., Barber, J., Cabanas, E. D., Collins, A., Donnai, D., ... & Pindar, L. (2005). 3q29 microdeletion syndrome: clinical and molecular characterization of a new syndrome. *The American Journal of Human Genetics*, *77*(1), 154-160.

Vega, W. A., Kolody, B., Aguilar-Gaxiola, S., Alderete, E., Catalano, R., & Caraveo-Anduaga, J. (1998). Lifetime prevalence of DSM-III-R psychiatric disorders among urban and rural Mexican Americans in California. *Archives of general psychiatry*, *55*(9), 771-778.

References

The Emory 3q29 Project. (2015, January 10). The Emory 3q29 Project. Retrieved September 21, 2018, from http://genome.emory.edu/3q29/

Mulle JG, Gambello MJ, Cook EH, et al. 3q29 Recurrent Deletion. 2016 Sep 22 [Updated 2017 Oct 19]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK385289/

SAS Institute Inc 2013. SAS/ACCESS® 9.4 Interface to ADABAS: Reference. Cary, NC: SAS Institute Inc.

SimplyAnalytics (2018). Census 2018 Current Estimates Data. Retrieved October 15th, 2018, from SimplyAnalytics database.

Ratcliffe, M., Burd, C., Holder, K., and Fields, A., (2016) "Defining Rural at the U.S. Census Bureau," ACSGEO-1, U.S. Census Bureau.

Tables and Figures

Gene	Function	Source
DLG1	Plays a role in trafficking of α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) and N-methyl-D-aspartate (NMDA) receptors to the neuronal membrane.	Mulle et al., 2017
FBXO45	Required for synapse formation, axon path finding, and neuronal migration	Mulle et al., 2017
PAK2	Involved in mediating molecular processes (neuron migration, outgrowth, and spine morphogenesis, and also controls neuronal differentiation)	Mulle et al., 2017
RNF168	Part of the cellular machinery responsible for the repair of DNA double-strand breaks.	Mulle et al., 2017
BDH	Plays a role in ketone body metabolism and regulation of lipid metabolism	Willat, 2005; Adami et al, 1993

Table 1. Select genes associated with 3q29 Deletion Syndrome

		All	Urban	Rural
Sex	Male	63	31	32
	Female	44	25	19
Autism (missing 5)	Yes	31	16	15
	No	73	38	35
Anxiety (missing 6)	Yes	30	13	17
	No	73	41	32
Failure to thrive (missing 6)	Yes	37	22	15
	No	66	32	34
Heart Defect (missing 2)	Yes	27	15	12
	No	80	41	39
Dental abnormality (missing 2)	Yes	91	45	46
	No	16	11	5
Age	*average*	13	9.93	15.87
Birth weight	*average*	6.22	6.17	6.28
Total		**107**	**56**	**51**

Table 2. Population Characteristics of Sample

	N	Mean	Standard Deviation	Variance	Missing
Anxiety					
Urban	54	24.07	0.43	0.17	2
Rural	49	34.69	0.48	0.23	2
Autism					
Urban	56	29.63	0.46	0.21	2
Rural	50	30	0.46	0.211	1
Failure to Thrive					
Urban	54	40.74	0.50	0.25	2
Rural	49	30.61	0.47	0.22	2
Heart Defect					
Urban	56	26.79	0.45	0.20	0
Rural	51	23.53	0.43	0.18	0
Dental Abnormality					
Urban	56	80.36	0.40	0.16	0
Rural	51	90.20	0.30	0.09	0

Table 3. Descriptive Characteristics for Dichotomous Variables

	N	Mean	Standard Deviation	Variance	Missing
Age					
Urban	56	9.93	6.00	35.98	0
Rural	50	15.87	13.46	181.12	1
Birthweight					
Urban	53	6.17	1.70	2.91	3
Rural	49	6.28	1.43	2.05	2
Distance to hospital (miles)					
Urban	56	5.98	4.76	22.69	0
Rural	51	18.30	31.86	1014.82	0

Table 4. Descriptive Characteristics for Continuous Variables

Variable	Method	Covariates	OR (CI)	Beta	p-value	P-value for covariates
Anxiety	Logistic regression	Sex, age	0.653 (0.267, 1.598)	-0.4257 (sex=0.0512, age =0.0162)	0.3511	Sex: 0.9085 Age: 0.4223
Autism	Logistic regression	Sex, age	1.074 (0.439, 2.626)	0.4563 (sex=0.4637, age= 0.0208)	0.8758	Sex: 0.0654 Age: 0.6377
Failure to thrive	Logistic regression	None	1.558 (0.69, 3.519)	0.4436	0.2859	
Dental abnormalities	Logistic regression	None	0.445 (0.143, 1.382)	-0.8104	0.1614	
Heart defect	Logistic regression	None	1.189 (0.495, 2.857)	0.1731	0.6987	
Birth weight	Linear regression	Gestational age, sex	NA	-0.11991 (gestational age= 0.30039, sex= -0.02519)	0.6319	Gestational age: <.0001 Sex: 0.9291

Table 5. Regressions with living in an urban area as the reference point.

Variable	Method	Covariates	OR (CI)	Beta	p-value	P-value for covariates
Anxiety	Logistic regression	Sex, age, distance to nearest hospital	0.596 (0.233, 1.521)	-0.5180	0.2786	Sex: 0.9296 Age: 0.4471 Distance to hospital: 0.5651
Autism	Logistic regression	Sex, age, distance to nearest hospital	1.174 (0.465, 2.962)	0.1605	0.7338	Sex: 0.0710 Age: 0.6167 Distance to hospital: 0.4642
Failure to thrive	Logistic regression	Distance to nearest hospital	1.527 (0.655, 3.563)	0.4235	0.3270	Distance to hospital: 0.8691
Dental abnormalities	Logistic regression	Distance to nearest hospital	0.589 (0.169, 2.044)	-5.300	0.4041	Distance to hospital: 0.3982
Heart defect	Logistic regression	Distance to nearest hospital	1.541 (0.588, 4.036)	0.4325	0.3786	Distance to hospital: 0.2549
Birth weight	Linear regression	Gestational age, sex , distance to nearest hospital	NA	-0.29454	0.2577	Gestational age: <.0001 Sex: 0.9604 Distance to hospital: 0.3593

Table 6. Regressions with living in an urban area as the reference point including distance to hospital as a covariate for all variables.

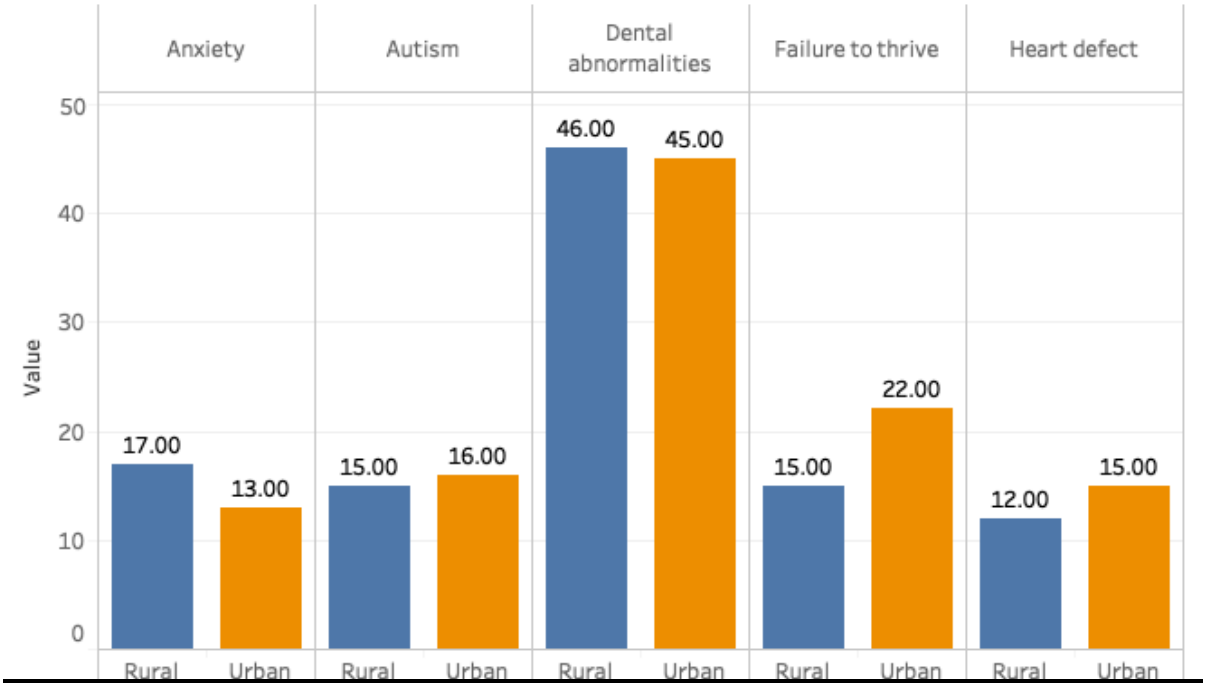

Figure 1. Counts of 3q29 associated phenotypes stratified by urban and rural populations.

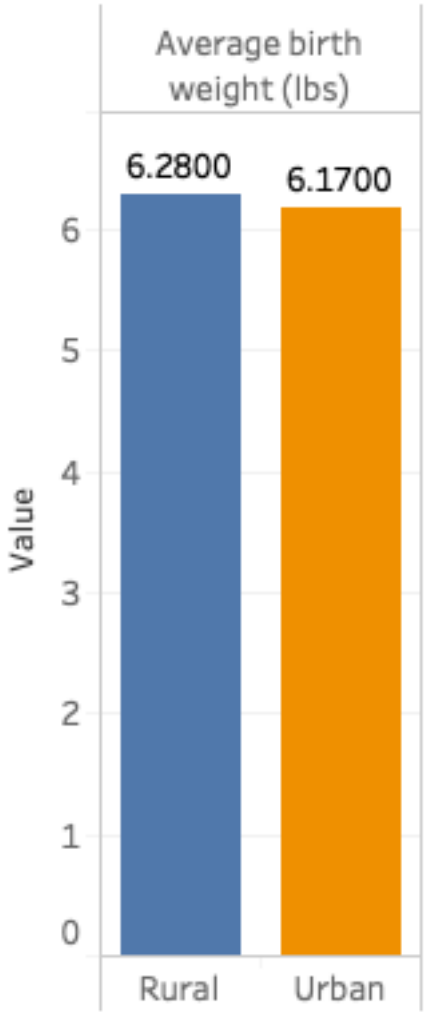

Figure 2. Mean birth weight of individuals with 3q29 Deletion Syndrome stratified by urban and rural living areas.

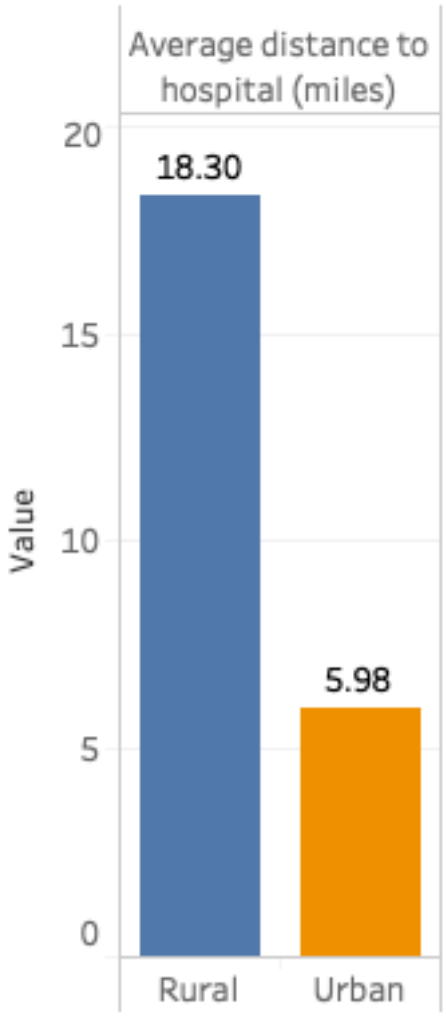

Figure 3. Average distance to hospital from reported zip code in miles for people with 3q29 Deletion Syndrome stratified by urban and rural living areas (p=0.0086).*