Long Run Health Effects of the Neolithic Revolution: The Natural Selection of Infectious Disease Resistance

Working Draft

C. Justin Cook*

January 15, 2013

Abstract

This paper argues that genetic differences arising since the Neolithic Revolution can explain variation in life expectancy before the advent of effective medicines and vaccines. The differential timing of the Neolithic Revolution, along with the varied distribution of domesticate animals, resulted in historical differences in the exposure to a class of infectious pathogens. This historical difference in disease environments led to differences in the natural selection of resistance. To measure the resulting genetic differences, we construct a measure of genetic variation within the human leukocyte antigen (HLA) system, a key component of the immune system. Our estimations show that a ten percent increase in HLA genetic diversity is roughly associated with a five percent increase in pre-medicinal life expectancy.

JEL Classification: N30, I12, Z13.

Keywords: Life Expectancy, Infectious Disease, Genetic Diversity, Neolithic Revolution, Natural Selection.

*Postdoctoral Associate, Division of Health Policy and Management,Yale School of Public Health;60 College St., New Haven, CT 06510; e-mail: charles.cook@yale.edu.

1 Introduction

The role of infectious diseases in low income economies is devastating in both a humanitarian and, arguably, an economic sense. In the developed world the majority of diseases are associated with aging, e.g. cancer and heart disease, but in developing states preventable infectious diseases remain the primary cause of mortality (Cutler et al. 2006). Why are these diseases so prevalent and destructive in developing states? The most obvious answer is tied to the socioeconomic conditions that continually plague inhabitants of less developed countries. The widespread incidence and mortality associated with infectious disease is seen as a byproduct of a lack of hygiene, insufficient preventative measures, and inadequate treatment options, which ultimately stem from low levels of economic development. This work questions whether there is a more fundamental determinant to the destructiveness of infectious disease. Particularly, are differences in disease resistance influenced by historical exposure? A notorious illustration is given by the numerous contacts between European explorers and native populations of the Americas and Oceania, in which previously unexposed native populations exhibited greater susceptibility and loss of life from many common European diseases.[1] With this idea in mind, the current work explores the historical origins of infectious pathogens in order to exploit the natural selection of resistance to these pathogens. Furthermore, we argue that these genetic differences have remained to the present and are associated with the efficaciousness of infectious disease before the widespread distribution of effective medicines and vaccines.

To test this idea we construct a measure of inherent genetic resistance for a group of infectious pathogens that developed as a result of the Neolithic Revolution. Differences in environments have led to differences in the initiation and sustainability of infectious diseases. This difference in disease environments led to differences in the natural selection of genes associated with disease resistance. The selection of resistance is based on variation within the genome rather than gene variants associated with particular traits; this is referred to as balancing selection (Jeffrey and Bangham 2000; King and Lively 2012; Prugnolle et

[1] Another example is given by the sickle cell trait, which provides resistance to malaria in Sub-Saharan Africa. The sickle cell trait is the result of historical exposure to malaria.

al. 2005).[2] In other words, the large number of diseases developing from agriculture, and the numerous strains associated with these diseases, resulted in the natural selection of variation *within* populations. With this idea in mind, we create a measure of genetic diversity that is based solely on gene variants found within a key component of the immune system, the major histocompatibility complex (MHC). In humans the MHC is referred to as the human leukocyte antigen (HLA) region, and this region contains the greatest amount of diversity within the human genome (Hughes and Yeager 1998; Jeffrey and Bangham 2000).[3] Our primary hypothesis explores the effect of these differences in HLA diversity on aggregate health outcomes before the widespread distribution and use of effective medicines and vaccines.

Our measure of genetic diversity, in theory, results from natural selection since the Neolithic Revolution. The effects of the agricultural transition on historic disease environments are twofold. First, agriculture allowed for the development of large, dense populations. Large, dense populations, in turn, allowed for an ease in the transmission of infectious diseases, as well as a large number of potential hosts. It is for this reason that diseases resulting from the Neolithic Revolution are known as crowd diseases (Wolfe et al. 2007). Second, the domestication of animals in the Neolithic provided closer contact between animals and humans. This close contact allowed animal pathogens to infect human hosts (Wolfe et al. 2007). The Neolithic Revolution provided the conditions for the initiation and sustainability of infectious crowd diseases. Societies which domesticated animals earlier and developed large, dense populations were the ones most likely to encounter infectious crowd diseases. This earlier exposure to disease led to the selection of genes that provide resistance to these diseases.

Starting with the Columbian Exchange, and accelerated by the mass development of roads, rail lines, and airports, infectious pathogens initiated by the Neolithic Revolution have spread throughout the world (Crosby 1973; Arroyo et al. 2006; Brownstein et al. 2006; Wilson 1995). This rapid pace of globalization has created a global disease pool, re-

[2]Section 2.2 gives a full explanation of balancing selection.

[3]MHC and HLA can be used interchangeably when discussing humans.

sulting in the introduction of diseases into previously unexposed populations (McNeil 1976; Harrison 2004). Given the slow change of the genome, genetic resistance, or susceptibility, is assumed to be relatively constant during this period of globalization, while exposure to infectious pathogens has changed. This implies that differences in selection due to historical exposure provide an inherent protection to this global disease environment, and should therefore be associated with differing health outcomes. We find that a 10 percent increase in HLA variation is associated with roughly a 5 to 6 percent increase in premedicinal life expectancy. This result is robust to numerous sample adjustments, the inclusion of potentially confounding variables, and is strengthened with the use of a previously discovered instrumentation strategy.

The rest of our paper is organized as follows. The next sub-section briefly outlines related literature. In Section 2 we discuss the initiation of infectious disease from the Neolithic Revolution and the resulting selection of genes within the HLA system. In Section 3 we describe the creation of our country-level measure of genetic diversity. In Section 4 we discuss additional data, our findings are discussed in Section 5, and Section 6 concludes.

1.1 Previous Literature

The role of historical environmental differences on contemporary health outcomes is explored in Galor and Moav (2007; hereafter GM). The authors show that an earlier transition to agriculture is associated with higher levels of contemporary life expectancy. Those societies who adopted agriculture at an earlier date have had an advantage in adapting to the new agricultural environment. In particular, GM (P. 1) state, "...the Neolithic transition altered the evolutionary optimal allocation of resources towards somatic investment, repairs, and maintenance (e.g., enhanced immune system, DNA repairs, accurate gene regulation, tumor suppression, and antioxidants)." To test this theory GM use a migration weighted average of the millennia a particular country has practiced agriculture. The authors find a strong positive relationship between a prolonged history of agriculture and variations in health outcomes in the year 2000, which gives credence to theory of adaptation proposed by GM. However, GM aren't able to measure adaptation directly; the weighted millennia

of agriculture is a proxy for adaptation. The current work seeks to build on that of GM by measuring a specific adaptation to the agricultural environment: genetic variation within the HLA region.

The relationship between aggregate genetic variations and economic outcomes is explored in a number of new papers. Ashraf and Galor (forthcoming) show that genetic variation within a country leads to differences in historical and contemporary levels of development. The authors suggest this is due to benefits resulting from moderate levels of genetic variation, where lower levels of genetic variation are associated with lessened creativity and higher levels of genetic variation are associated with frequent in-fighting. Our method of measuring genetic diversity is identical to that found in Ashraf and Galor (forthcoming); however, we consider genes within the HLA system that have been influenced by presence of infectious pathogens, not overall genetic differences. A complete discussion of this variable is given in Section 3.1. Additionally, our instrumentation strategy is based on the work of Ashraf and Galor (forthcoming).

Spolaore and Wacziarg (2008; hereafter SW) use genetic difference as an explanation for the diffusion of technology between states. In short, SW theorize that a greater genetic distance to the technological frontier is associated with a lag in the adoption of new technologies and, therefore, a lower level of income.[4] SW consider only random, or neutral, mutations that have occurred due to the separation between populations; we, on the other hand, consider genes which have encountered selection.[5] In addition to the work of SW, Guiso et al. (2004) and Giuliano et al. (2006) use genetic distance in the explanation of bilateral trade flows between countries in Europe.

Finally, a number of theoretical works discuss historical adaptation in regards to contemporary variations in wealth and health. These include, but are not limited to: Galor (2011), Galor and Moav (2007), and Galor and Weil (1999).

[4]The technological frontier is the U.S. in the year 2000 and Great Britain in the year 1500. The measure of genetic distance which is used in SW is Sewall Wright's fixation index, or a measure of genetic diversity *between* populations.

[5]To test for directional selection, or the selection to unity of a gene that generates a favorable trait, we construct a measure of genetic distance comprised of HLA genes. This is discussed in Section 5.3.

2 The Neolithic Revolution and the Natural Selection of Disease Resistance

This section will i.) provide further evidence for the role of the Neolithic Revolution in the initiation and sustainability of infectious crowd diseases and ii.) describe selection for variation within the HLA region.

2.1 Crowd Disease

The rise of infectious disease in man is dependent upon agriculture. The domestication of animals created close contact between farmers and their animals, which allowed animal pathogens to infect new human hosts (Wolfe et al. 2007). Diphtheria, influenza A, measles, mumps, pertussis (whooping cough), rotavirus A, smallpox, and tuberculosis are similar to pathogens afflicting domesticate animals of Eurasia and "probably or possibly reached humans from domesticate animals (Wolfe et al. 2007, P. 281)." Those peoples who first domesticated a particular animal had a greater probability of contracting the domesticate's diseases; implying that the peoples of Eurasia, with the highest number of potential domesticate animals, were the initial hosts of many crowd diseases (Diamond 1998; Hibbs and Olsson 2004).

In addition to initiation, the sustainability of a pathogen within a population is necessary for the development of genetic resistance. Endemicity, or the sustained presence, of infectious crowd disease is dependent on population density (Dobson 1996; Anderson and May 1991; Wolfe et al. 2007). In order for a disease to persist within a population, the population must be large enough so that newly susceptible individuals, or hosts, are present. The diseases of interest in this paper are those that either kill the host or provide the host with antibodies so that he or she develops immunity to the disease. This implies that in small populations all susceptible individuals will either die or become immune, causing the disease itself to die out. As an example, it has been shown that measles becomes endemic in island communities with populations roughly greater than 500,000 individuals (Black 1966). If populations aren't sufficient in size, epidemics occur in which the disease sweeps through a population, leaving its members either dead or immune.

Hunter-gather societies could not support large enough populations for endemic diseases. Only the relatively large societies resulting from agriculture can supply hosts in such large numbers in which the disease could be continually maintained. Eurasian countries contained an advantage in the initiation of agriculture, implying these states had the necessary population size to replenish hosts necessary for the endemicity of the pathogens under consideration (Diamond 1998). Larger populations also led to greater cities that facilitated the spread of disease through closer contacts and lower hygiene (McNeil 1976). Additionally, the sedentary lifestyle of the agricultural environment allowed for the contamination of water supplies and the collection of rodents and other pests that carry vectors for disease.[6]

A hypothesis posed by Barnes et al. (2010) corroborates the role of dense populations in selection for resistance. Barnes et al. (2010) find that a history of living within a city has a close association with genetic resistance to tuberculosis. In summary, the dense populations associated with cities have allowed for a greater spread and sustainability of tuberculosis, which in turn, has led to a greater selection for an allele that provides resistance to tuberculosis.[7] Their findings suggest differential disease environments have led to differences in adaptation. This idea is naturally extended by exploring the effect of differential adaptation on contemporary health outcomes; this is our main hypothesis.

While large populations are necessary for the endemicity of infectious diseases, they are not sufficient in explaining the differences in disease resistant alleles. This point is most apparent when considering the ruinous results disease played on New World populations. The Mayans, Aztecs, Incas, and certain North Amerindian communities all developed agriculture and had populations sufficient in size to support the endemicity of disease, yet obviously had not developed an inherent genetic resistance to sustain the diseases from European conquerors and settlers during the colonial period (Crosby 1986; Mann 2005). The reason is that these societies simply had no exposure to a large number of infectious pathogens; if

[6]Plague and typhus are primarily distributed through lice, which are native to rodents that can only be supported in large sedentary human settlements.

[7]Selection for resistance to tuberculosis, relative to other crowd diseases, should occur at a slower rate. Tuberculosis has a golden age of about 15-25 in which the mortality of the disease becomes quite low; this corresponds to the ability to produce offspring. Measles and mumps, on the other hand, usually affect infants who are unable to produce offspring.

the selective force (i.e., disease) is not present, then selection for disease resistance will not take place.

The development of infectious crowd disease is dependent upon both the wide domestication of animals and large, dense societies. Eurasia contained the advantage of contracting diseases earlier and also having large enough populations in which to sustain the particular diseases. In the words of Wolfe et al. (2007, P. 281):

> Thus, the rise of agriculture starting 11,000 years ago played multiple roles in the evolution of animal pathogens into human pathogens. Those roles included both generation of the large human populations necessary for the evolution and persistence of human crowd diseases, and generation of large populations of domestic animals. Moreover, as illustrated by influenza A, these domestic animal herds served as efficient conduits for pathogen transfers from wild animals to humans, and in the process may have evolved specialized crowd diseases of their own.

This process led to a continual selection for individuals containing a greater inherent resistance. The selection process is explored in the next section.

2.2 Pathogen Driven Selection

If a disease enters into a primitive society in which no medicine exists, some individuals may die from the disease while others may not. It is this variation amongst individuals, corresponding to variations within the genome, which causes disease resistance to be selected. This is natural selection in which the strong survive, where in this case, strength is determined by some unseen phenotypic difference that allows some to be more resistant to disease (e.g., better recognition of potential infections, better disposal of harmful pathogens, etc.).[8] Furthermore, disease environments have differed, leading to differences in selection. This concept is expressed by Inhorn and Brown (1990, P. 89):

[8] A phenotype is the expression of the genotype (Hartl and Clark 2007). The recognition and response of the immune system is a phenotypic expression of underlying genes.

> . . . infectious diseases including both great epidemics, such as plague and small pox, which have devastated human populations from ancient to modern times, and less dramatic, unnamed viral and bacterial infections causing high infant mortality have likely claimed more lives than all wars, noninfectious diseases, and natural disasters taken together. In the face of such attack by microscopic invaders, human populations have been forced to adapt to infectious agents on the levels of both genes and culture.[9] As agents of natural selection, infectious diseases have played a major role in the evolution of the human species.

Therefore, holding socioeconomic conditions constant across peoples, those societies that have been in contact with infectious crowd diseases for longer periods, or have had more time to adapt to the infectious pathogens, should contain greater genetic resistance to these particular diseases.

The high level of variation within the HLA system is based on a theory of balanced selection (de Bakker et al. 2006; Jeffrey and Bangham 2000; Traherne et al. 2006; Klein 1987). Balancing selection is selection for genetic diversity and results from two distinct reasons: overdominance and frequency-dependence (Slade and McCallum 1992). Overdominance implies heterozygotes, or individuals with differing alleles at a particular locus, have an advantage compared to homozygotes, or individuals with identical alleles at a particular locus.[10] A prime example of overdominance is the advantage conferred by the sickle-cell trait (Allison 1956). Heterozygous individuals contain a greater resistance to malaria, while homozygotes either contain no resistance to malaria or are afflicted by sickle-cell anemia. This leads to the natural selection of variation at the gene locus responsible for the the sickle-cell trait.

Frequency-dependent selection results from a comparative advantage of rare alleles. Infectious pathogens don't constitute a static selection pressure. Infectious pathogens—bacteria, viruses, protozoa, etc.—are living things also undergoing natural selection. If a

[9] An example of cultural adaptation would be the washing hands, thoroughly cooking meat, or the wearing of a surgical mask while in public. Footnote our own.

[10] Individuals contain alleles at a gene locus from both the mother and father, implying two alleles at a given locus.

particular allele were to provide complete resistance to a certain pathogen, variants of the pathogen, which avoid resistance, would thrive. In effect, this results in an "arms race" between the pathogen and the person. The relatively short time between generations of most pathogens, however, provides a time advantage in this "arms race." This implies that any resistance developed in the human genome should be overcome by genetic mutations within the pathogen. In other words, infectious pathogens have greater defenses to more common HLA gene variants; therefore, rarer, or lesser frequent, HLA alleles are better able to recognize and dispose of disease, implying a constant selection for rarer HLA alleles. As a result, the optimal strategy for disease resistance is variation, or allowing alleles associated with recognition to be played in equal frequency. Prugnolle et al. (2005) confirm this idea in finding that pathogen richness, or a high number of infectious pathogens, is associated with diversity within the HLA system. Furthermore, adaptation of infectious pathogens is routinely seen in the development of antibiotic resistance.

Natural selection has taken place within the HLA system since the initiation of agriculture (Sabetti et al. 2006). This natural selection has led to high levels of diversity within the HLA system, implying balancing selection (Prugnolle et al. 2005). Differences in HLA diversity remain, and these differences affect the immune response to the large number of diseases developed during the Neolithic Revolution (Bhatia et al. 1995; Black 1994; Black et al. 1974). Given differences in immune response and the widespread distribution of Neolithic crowd diseases, we postulate that differences in HLA diversity have an affect on contemporary health outcomes before the distribution of effective medicines and vaccines.

3 Disease Based Genetic Diversity

This section will outline the creation of the genetic diversity measure. First, a commonly used measure of genetic diversity is described. Second, we will discuss specific gene variants that will comprise our measure of genetic diversity. Finally, we will discuss aggregation to the country level.

3.1 A Measure of Genetic Diversity

In a recent work Ashraf and Galor (forthcoming; hereafter AG) explore the role of genetic variation in explaining historical and contemporary levels of development. In order to measure genetic diversity, AG use a common measure within population genetics: expected heterozygosity. Expected heterozygosity is roughly defined as "the probability that two randomly selected individuals differ with respect to the gene in question (AG, P. 3)." Expected heterozygosity is calculated with the frequency of gene variants, or alleles, at a particular site on the genome, or locus. Mathematically, expected heterozygosity is defined by:

$$H_{exp} = 1 - \frac{1}{m}\sum_{l=1}^{m}\sum_{i=1}^{k_l} p_i^2 \tag{1}$$

where p_i represents the fraction of allele i, and expected heterozygosity is found by the average across m loci.[11]

Our measure of heterozygosity differs slightly from that found in AG. AG attempt to measure variation within the entire genome in order to measure the effects of fractionalization and creativity associated with high and low levels of diversity, respectively. Our work differs in that we seek to measure heterozygosity in order to show balancing selection from the numerous infectious pathogens that became endemic after the Neolithic Revolution. Therefore, we only consider genes within a key component of the immune system, the major histocompatiblity complex.

3.2 Alleles Associated with Infectious Disease: The Major Histocompatibility Complex

The specific genes to be considered in the construction of our genetic diversity measure are based on the major histocompatibility complex . The major histocompatibility complex is a group of genes associated with the recognition of foreign substances within the body and is very important in disease resistance and susceptibility (Klein 1987; Traherne et al. 2006).

[11] An allele is a gene variant. As described in the next sub-section, we use 156 differing loci, which can take one of two possible values. This implies that expected heterozygosity is maximized when each $p_i = 0.5$.

In short, the MHC is responsible for locating foreign proteins in order to direct cells of the immune system to initiate an immune response (Piertney and Oliver 2006).

In humans the MHC is known as human leukocyte antigen system (Encyclopedia Britannica 2011).[12] The HLA system is a cluster of 239 genes located on the sixth chromosome (Shiina et al. 2004). The MHC is broken into two major classes, Class I and Class II, with both classes being associated with the recognition of certain pathogens.[13] This work, however, targets the entire system, and not sole genes. The use of all gene variants within the HLA system allows for a more complete measurement of diversity resulting from exposure the numerous Neolithic crowd diseases.

In the construction of our expected heterozygosity measure we consider gene variants within the HLA system; therefore, our main measure is HLA heterozygosity. HLA heterozygosity is constructed with data on SNP's from the Allele Frequency Database at Yale University, referred to as ALFRED. A SNP (pronounced "snip") is a single change along a strand of DNA. From the website, "ALFRED is a free, web-accessible, curated compilation of allele frequency data on DNA sequence polymorphisms in anthropologically defined human populations." The use of the ALFRED gives allele frequency data for 156 SNPs of 19 HLA genes for 51 ethnic groups. Using Equation (1), the 156 SNPs constitute the m loci. Each SNP has two variants, or alleles; therefore, expected heterozygosity for a single locus is maxed when the two possible variants of a SNP equal 50%. HLA heterozygosity is the average for all HLA SNPs.

The HLA system is associated with resistance and susceptibility to infectious disease (Traherne et al. 2006). Theoretical differences should exist within this system due to differences in the disease environments (Jeffrey and Bangham 2000). Our primary measure quantifies these differences by measuring diversity within the HLA system. The next subsection explains the aggregation of ethnic groups to the countries.

[12]White blood cells are known as leukocytes. MHC and HLA can be used interchangeably when discussing humans.

[13]In the recognition of cells, Class I molecules are expressed on nucleated cells and are associated with defense against viruses, while Class II molecules are expressed on antigen-presenting cells and are associated with extracellular parasites (Piertney and Oliver 2006).

3.3 Aggregation from Ethnic Groups to Country

Allele frequency data is given by distinct ethnic groups; however, many (or most) relevant economic data are at the country level. This implies that an aggregation is needed in which countries are constructed of ethnic groups. Following Spolaore and Wacziarg (2009), we aggregate ethnic groups to the country level with the use of ethnic compositions found in Alesina et al. (2003). [14]

The matching of ethnic groups from ALFRED to Alesina et al. (2003) is not perfect. ALFRED contains allele frequency data for 51 differing ethnic groups, while Alesina et al. (2003; hereafter Alesina) contains hundreds of differing ethnic groups. In order to get around this problem, language classifications are used to match distinct ethnic groups in Alesina to a similar ethnic group in ALFRED (Lewis 2009). For example, Hutu from Alesina are classified as Bantu in Alfred, Amayara are classified as Amerindian, and Polish are classified as Russian.

In addition to the matching of ethnic groups, additional ethnic groups have been created through combinations of ethnicities found in ALFRED. [15] The primary example of this is given by the ethnicity Black in Alesina et al. (2003). The term Black refers only to the color of skin, not ethnicity. Ultimately, Black indicates a hereditary history from Sub-Saharan Africa, but Sub-Saharan Africa is not made up of a sole ethnic group. In order to get around this problem, we first assign Sub-Saharan African countries to one of three ethnic groups based on a map in Shillington (1989, P. 50; Reader 2002, P. 692)[16] Next, using data on the Trans-Atlantic slave trade from Nunn (2009), we create a representative Black ethnic group through the weighted average of the number of slaves from an African country that has been assigned to a specific ethnic group. This leads to the representative Black ethnic group comprised of 49% Bantu, 12% Mandenka, and 39% Yoruba. Other notable

[14] The ethnic compositions found in Alesina et al. (2003) are from the 1990's. This creates an error in measurement for HLA heterozygosity. However, there is no reason to suspect a nonrandom error. Therefore, this measurement should lead to an attenuation bias, understating the true relationship of HLA genetic distance. This issue is addressed with use of an exogenous instrument.

[15] These combinations are not counted in the calculation of the heterozygosity score.

[16] West African countries are assigned to Mandenka, countries around the Gulf of Guinea are assigned to Yoruba, and South African countries are assigned to Bantu. Note that most Northeast African/Nilo-Saharan states are unused due to the lack of a close ethnic group in ALFRED.

combinations include: White which is 50% Italian and 50% French, Mestizo which is 50% White and 50% Amerindian or Mayan (depending on whether the respective country is in North or South America), and Germanic which is 50% French and 50% Orcadian. Through this method I am able to find the genetic diversity for 175 countries, of which 124 are used in our baseline regression model. A global plot of cross-country HLA heterozygosity is given in Figure 1; note that countries with populations derived from Eurasia tend to have greater levels of HLA diversity. Sub-Saharan African countries also contain high variation within the HLA system; this is explained in the next section.

4 Other Data

Table 1 gives summary statistics for all variables used in the baseline estimation. The origin of our measure of HLA heterozygosity is given in detail above, while sources and explanations of all control variables are given in Appendix A. Below, we describe the reasoning of our main dependent variable, as well as our primary instrument for 2SLS estimation.

4.1 Dependent Variable

The primary health variable to be considered in this paper is life expectancy in 1960 (WDI). The use of 1960 life expectancy is meant to capture health variations before the widespread distribution of effective medicines and vaccines (Acemoglu and Johnson 2006).[17] In theory, our measure of genetic diversity should affect the resistance of a country's population to infectious crowd disease in the absence of medicine; i.e., if a country's population has relatively low levels HLA heterozygosity, then in the absence of medicine, a greater fraction of the population will die from infectious crowd diseases. This higher mortality then is associated with a lower level of life expectancy; additionally, the diseases under consideration disproportionately affect infants and young children, which further affects life expectancy. Life expectancy indirectly measures the burden of disease, and this burden of disease is

[17] Acemoglu and Johnson (2006) exploit the "epidemiological transition" which began in the 1940s. While many vaccines and medicines were invented in the 1940s and 1950s, the widespread distribution and use of these medicines was slowed. Earlier years are considered, but due to a lack of measurement in relatively poor countries, very few data are available. Therefore, the use of 1960 is seen as a trade-off between data and the timing of medicinal distributions.

more randomly assigned before the widespread use of effective medicines and vaccines.

4.2 Migratory Distance from East Africa

Homo sapians originated within Africa roughly 200,000 years ago; around 100,000 years ago modern humans began to migrate out of East Africa, resulting in human colonization of the entire planet (Ashraf and Galor 2012; Prugnolle et al. 2005a; Ramachandran et al. 2005). This process of migrating out of Africa resulted in population bottlenecks and a decline in genetic diversity as a result of migratory distance. In other words, migrating populations carried only a fraction of genetic diversity, reducing the overall level of genetic diversity within the sub-population. This implies a clear linear relationship between heterozygosity and migratory distance from East Africa, the jumping off point for the "Out of Africa" migrations.

This relationship is exploited in Ashraf and Galor (forthcoming). Figure 2 plots the expected migratory paths out of East Africa, in which "Out of Africa" migratory distance is calculated as the sum of the distance between country and the closest way point and the distance of this way point to East Africa (along the proposed migratory path). Using neutral (not solely HLA) genetic variation, the migratory distance from East Africa explains roughly 85% of the variation in expected heterozygosity. Given this strong relationship, AG use migratory distance to predict country level heterozygosity.

Our measure of HLA heterozygosity isn't based on neutral variation; it is based on a key component of the immune system that has under gone recent selection (Sabettie et al. 2006). This strong linear relationship between "Out of Africa" migratory distance and HLA heterozygosity need not be the case, since genetic bottlenecks due to the serial founder effect are not the only factor in determining HLA heterozygosity. Prugnolle et al. (2005b) find migratory distance explains only 17% to 39% of diversity within HLA genes, not 86% as found in Ashraf and Galor (2012). In explaining the residual variation in HLA heterozygosity, Prugnolle et al. (2005b) show that pathogen richness has a strong, positive association with variation within HLA genes.[18] This implies infectious pathogens

[18]Pathogen richness is the total number of intracellular diseases within a country.

are responsible for shaping the diversity within HLA genes.

Instead of contemporary pathogen richness, which is associated with our dependent variable, we too use migratory distance from East Africa as an instrument for HLA heterozygosity; however, we exploit the nonlinear relationship that migratory distance has with HLA heterozygosity. This nonlinear relationship is the result of disease initiation and duration within Eurasia, from which HLA heterozygosity increases for populations that moved into Eurasia but then falls as migratory distance increases. Graphically this is shown in Figure 3. African countries have a relatively high HLA heterozygosity and a short migratory distance from East Africa. European, neo-European, and Middle Eastern states, however, have the greatest level of HLA heterozygosity, which is consistent with the earlier exposure to domesticate animals and denser populations associated with populations of these regions. Furthermore, HLA Heterozygosity falls rapidly as migratory distance is increased for American and Oceanic states. We therefore use migratory distance and its square as instruments for estimating the effect of HLA heterozygosity on premedicinal life expectancy. The use of migratory distance as an instrument is primarily done to correct for measurement error that results from aggregating ethnic data to the country level.

5 Results

5.1 Explaining HLA Heterozygosity

Two primary factors are responsible for explaining heterozygosity within the HLA system. First, heterozygosity is a declining function of the distance from East Africa. Due to the serial founder effect, genetic variation within populations declines as the migratory distance from East Africa increases. Due to the infectious disease presence in Eurasia, however, a nonlinear relationship exists between the out-of-Africa migratory distance and HLA heterozygosity. This nonlinearity is shown by an initial increase in HLA heterozygosity due to out-of-Africa migratory distance, which is then followed by a steady decline; this is shown in Figure 3. Second, the early development of agriculture and the intense domestication of animals within Eurasia facilitated the development and duration of a large number of infectious pathogens. These factors differed across states and peoples, implying differences

in HLA heterozygosity. Table 2 displays the estimated relationship for these two proposed determinants of HLA heterozygosity.

The relationship between "Out of Africa" migratory distance and heterozygosity within the HLA system is explored in column (1) of Table 2. Column (1) regresses HLA heterozygosity on the migratory distance from East Africa and its square; both coefficients are significant at the 1% level and have a joint F statistic of 69.72. The coefficients on migratory distance and its square indicate HLA heterozygosity increased up to a certain distance outside of Africa before decreasing due to the serial founder effect, with maximal heterozygosity found 4,532 km from East Africa. This distance corresponds to countries within the Middle East and Southern Europe, and further verifies the role of agriculture in shaping variation within the immune system.

In addition to migratory distance from East Africa, the historical environment of Eurasia also influenced HLA heterozygosity. The Neolithic Revolution provided the means to develop large, dense populations, but a dense population is the factor that allowed for the persistence and spread of infectious crowd disease. To confirm this idea, we use population density in 1 CE as an explanatory factor of HLA heterozygosity (McEvedy and Jones 1978). The bivariate relationship between HLA heterozygosity and historical population density is found in column (2) of Table 2. Historic population density has a positive and statistically significant relationship with HLA heterozygosity; the coefficient of column (1) implies that as population density increases by 10%, HLA heterozygosity increases by 0.2%. For Poland, the country with the median population density in 1 CE, an increase of one person per squared kilometer would be associated with a 1.2% increase in HLA heterozygosity. Additionally, the initiation of the infectious pathogens under consideration is a function of close contacts between man and domesticate animals (Woolfe et al. 2007). We therefore use the number of potential domesticate animals from Hibbs and Olsson (2004) as a determinant of HLA heterozygosity. Column (3) of Table 2 shows that the number of potential domesticates is positively and statistically associated with our measure of HLA heterozygosity, where an increase of an additional potential domesticate animal is associated with roughly a one percent increase in HLA heterozygosity.

Column (4) includes both historic determinants of HLA heterozygosity as well as an interaction term. Both variables and the interaction are significant at the 1% level, although the point estimate on historic population density is negative. The marginal effect of population density in 1 CE is positive, however, for countries above the median number of potential domesticate animals (i.e., 7 potential domesticates). Importantly, the interaction of the two variables is positive, implying a complementarity in HLA heterozygosity. In other words, having both a dense population and a large number of domesticate animals is associated with increased HLA heterozygosity. The estimates of columns (2)-(4) indicate that the historical environment is associated with contemporary differences in HLA heterozygosity.

HLA diversity is a function of both distance from East Africa and the Neolithic Revolution.[19] Distance out of Africa, is associated with a decline in genetic diversity (Ramachandran et al. 2005). This loss in diversity, however, has been overturned within Eurasia. This is due primarily to the development of a large array of diseases from the initiation of agriculture. The Neolithic Revolution provided close contact between numerous domesticate animal species and humans. This close contact facilitated the transmission of novel pathogens into human populations. This transmission was supported by the large, sedentary populations that resulted from the Neolithic Revolution. The prolonged exposure associated with agriculture led to balancing selection for genes responsible in the recognition of foreign pathogens. Therefore, we should see greater variations within the HLA system for Eurasians. The estimates of Table 2 confirm these two primary effects on HLA heterozygosity. The next subsection explores whether this inherent genetic variation can explain differences in contemporary health differences.

5.2 HLA Heterozygosity and Premedicinal Health Outcomes

This subsection provides results for tests of our main hypothesis: HLA variation provided an inherent resistance to the numerous, Eurasian crowd diseases that were spread across the globe following European colonization. The spread of disease was further accelerated

[19]Due to the larger sample and larger explanatory power we use migratory distance as our main instrument. Additionally, historic population density and the number of potentially domesticated animals may have long run associations with wealth, and therefore do not serve as valid instruments.

by the widespread development of ports, roads, rail lines, and airports (Arroyo et al. 2006; Brownstein et al. 2006; Wilson 1995). Given the relatively slow pace of natural selection and the widespread distribution of infectious pathogens, we should expect differences in HLA heterozygosity to persist into contemporary times.

5.2.1 Health Prior to the Epidemiological Transition

A large number of effective medicines and vaccines were discovered in the late 1940's and early 1950's; this is referred to as the epidemiological transition. Acemoglu and Johnson (2007; hereafter AJ) exploit this exogenous discovery date to explore the effect increases in life expectancy have on economic growth. We too are focused on health outcomes prior to the widespread use of effective medicines and vaccines, so the initial periods considered in AJ serve as a valid starting point. Table 3 considers the effect of HLA heterozygosity on both life expectancy in 1940 and the predicted mortality from infectious disease in the same period; the two dependent variables constitute important health outcomes prior to the epidemiological transition.

Column (1) shows the bivariate relationship between HLA heterozygosity and predicted mortality from a large number of infectious diseases in 1940. The coefficient is negative and statistically significant at the 1% level, indicating greater variation within the HLA system is associated with lower mortality from infectious disease in 1940. In particular, a one standard deviation increase in HLA heterozygosity for the mean country in our sample would be associated with a 35% decline in mortality rates. Column (2) controls for income in 1940. This leads to an attenuation of the coefficient of HLA heterozygosity, but the effect of HLA heterozygosity remains both negative and statistically significant. Column (3) performs 2SLS estimation with migratory distance from East Africa and its square as the excluded instrument. The 2SLS estimates of column (3) are similar in both magnitude and significance to the OLS estimates of column (2). Columns (4)-(6) perform estimates similar to those of columns (1)-(3) with the exception being life expectancy in 1940 as the dependent variable. Columns (4)-(6) show HLA heterozygosity has a positive and strong statistical relationship with life expectancy before the epidemiological transition. The

estimated coefficient of interest in column (6) implies a greater than one-to-one relationship between HLA heterozygosity and life expectance in 1940; for the mean HLA heterozygosity, a one standard deviation increase is associate with roughly a 10% increase in life expectancy.

The estimates of Table 3 corroborate our main hypothesis: In the absence of medicine, peoples that evolved greater variation within the HLA system contain greater defenses against a large number of infectious pathogens. These greater defenses resulted in lower a rate of mortality from infectious disease and led to higher life expectancies.

5.2.2 Baseline Results

The use of 1940's health data is ideal; however, data are relatively sparse for this period. Important control variables are absent for this period, and accurate country-level data are restricted to the more developed nations of the time. We therefore sacrifice the potential benefits of using a time period before the epidemiological transition in order to have a more general sample.[20] Our main dependent variable is life expectancy in 1960. Given that many medicines and vaccines were discovered in the late 1940's and early 1950's, the use of 1960 data is meant to capture the effects of inherent resistance before the *widespread distribution* of these medicines and vaccines.[21]

Our baseline estimating equation is given by:

$$ln\ \mathrm{LE}_i^{1960} = \alpha + \beta_1(ln\ \mathrm{HLA}_i) + \boldsymbol{\beta_2'SE_i} + \boldsymbol{\beta_3'G_i} + \boldsymbol{\beta_4'I_i^c} + \epsilon_i$$

Where i is a country indicator, LE^{1960} is life expectancy in 1960, and HLA is our measure of HLA heterozygosity with β_1 the coefficient of interest throughout the paper. $\boldsymbol{SE_i}$ is a vector of country-level socioeconomic controls, including GDP per capita in 1960, years of schooling in 1960, and ethnic fractionalization; $\boldsymbol{G_i}$ is a vector of geographic controls, which includes absolute latitude and the fraction of a country within the tropics; $\boldsymbol{I_i^c}$ is an indicator variable as to whether or not country i is within continent c; and ϵ_i is the cross-country error term.

[20] Any potential effects of medicine should be captured within GDP per capita.

[21] This is checked in Sec. 5.3.1.

Column (1) of Table 4 gives the bivariate regression of 1960 life expectancy on HLA heterozygosity. The coefficient on HLA heterozygosity is positive and significant at the 1% level. Greater genetic diversity within the HLA system is associated with higher premedicinal life expectancy. Specifically, a 10% increase in HLA heterozygosity is associated with a 12.5% increase in life expectancy. Sub-Saharan African states contain a relatively high level HLA heterozygyosity with little variation between states; the small amount of HLA variation between Sub-Saharan African states is due to the limited number of ethnic groups used in calculating country-level measures for these states.[22] The low variation between states corresponds to high differences in life expectancy, implying HLA heterozygosity has little explanatory power for Sub-Saharan African states. In order to correct for this and to account for unobserved differences between continents, column (2) includes continent fixed effects into the bivariate regression of column (1). The inclusion of continental dummies attenuates the magnitude but does not affect the significance of the the coefficient of HLA heterozygosity.[23].

Column (3) includes the vector of socioeconomic controls into the estimation of column (2); these include GDP per capita in 1960, total years of schooling in 1960, and ethnic fractionalization. Health and income are strongly correlated; more developed states are better able to provide greater nutrition, sanitation, and care to the sick (Bloom and Canning 2000; Pritchett and Summers 1996). It is therefore necessary to control for income differentials in our baseline model. Additionally, we include years of schooling from Barro and Lee (2011) in our vector of socioeconomic controls due to the strong relationship between human capital, particularly education, and health (Baker et al. 2011; Kenkel 1991). Finally, in order to control for any unseen association between ethnic compositions in Alesina et al. (2003) and our dependent variable, we include ethnic fractionalization into our baseline socioeconomic controls.

The inclusion of socioeconomic controls in column (3) does lead to an attenuation of the coefficient of HLA heterzgosity; however, the coefficient remains positive and statistically

[22]Most countries within Sub-Saharan Africa are assigned to the Bantu ethnic group.

[23]Continental dummies include indicator variables for the Europe, Africa, Asia, North America, and South America

significant at the 1% level. The estimates of column (3) suggest that a 10% increase in the portion of the genome associated with the recognition of foreign pathogens is associated with a 5% increase in premedicinal life expectancy. Consider two countries with similar income levels in 1960: Venezuela and France. If Venezuela's population contained HLA variation similar to the population of France, Venezuela's life expectancy would increase by roughly 5 years, which would half the difference in life expectancy between the two countries.

Column (4) includes a vector of geographic controls into the bivariate regression with continent fixed effects found in column (2). Our baseline geographic controls include absolute latitude and the fraction of a country found within the tropics, which are used to capture the effects of contemporary environmental differences. The inclusion of geographic controls does not alter the coefficient of interest, which remains positive, significant, and similar in magnitude to the estimate of column (2).

All baseline controls are included in column (5), representing our baseline estimating equation. The effect of HLA heterozygosity is roughly half of the estimate given by the bivariate regression but is similar in magnitude when controlling for socioeconomic conditions. The effect of HLA heterozygosity is positive and significant at the 1% level for the baseline model. The estimated effect provides strong support for the main hypothesis of this paper: long running genetic differences within the immune system have an effect on health outcomes before the widespread distribution of medicine. The initiation of agriculture and the domestication of animals resulted in the development and sustainability of infectious pathogens. Prolonged exposure to these pathogens provided selection pressures favoring variation within the portion of the genome responsible for recognition of foreign bodies. Globalization resulted in the spread of these disease to populations that had no previous exposure, and therefore, lower diversity within the HLA system. Before efficacious medicines and vaccines, the difference in HLA diversity led to differences in how populations were able to cope with these infectious diseases, resulting in differences in life expectancy. This is shown in Table 4.

While the coefficient of HLA heterozygosity is statistically significant in all estimations of Table 4, the reported effect is likely being attenuated due to numerous errors in mea-

surement associated with aggregating an ethnic measure to a country-level measure. To account for this, we use instrumental variables. As shown in Ashraf and Galor (forthcoming), "Out of Africa" migratory distance is a strong predictor of heterozygosity, or variation within the genome. Given its exogeneity and natural relationship to genetic variation, the distance from East Africa serves as an ideal instrument for our measure of HLA heterozygosity. Therefore, as in column (1) of Table 2, we use migratory distance and its square as instrumental variables in Table 5.

The estimations of Table 5 mirror those of Table 4, with the 2SLS estimator being used in place of OLS. The use of migratory distance and its square are strong instruments for HLA heterozygosity. This is seen in the large first stage F statisitcs in Table 5, which average roughly 70. Additionally, we cannot reject the validity of migratory distance and its square as excluded instruments for the baseline estimating equation.

Column (1) gives the bivariate 2SLS estimates of regressing life expectancy in 1960 on HLA heterozygosity. The point estimate of the coefficient is positive but statistically insignificant. As stated earlier, Sub-Saharan African states have similar levels of HLA heterozygosity with a wide range of life expectancy in 1960. The inclusion of continent fixed effects in column (2) results in a statistically significant relationship between HLA heterozygosity and premedicinal life expectancy. Columns (3) and (4) include socioeconomic and geographic controls, respectively.

Column (5) of Table 5 satisfies the baseline estimating equation while instrumenting HLA heterozygosity with the migratory distance from East Africa and its square. The coefficient of HLA heterozygosity is positive, significant at the 1% level, and larger than one standard deviation of the estimate given by OLS. In particular, the estimated coefficient of column (7) states that a 10% increase in HLA heterozygosity leads to an 6.7% increase in life expectancy. For mean life expectancy in 1960, this corresponds to an increase in life expectancy of 4 years. For our baseline model, the use of migratory distance as an instrument results in a significantly larger coefficient and alleviates attenuation due to measurement error. Additionally, the use of an exogenous instrument also alleviates concerns for reverse causality and simultaneity bias. The results of Table 5 provide further support for our main

hypothesis.

Through both instrumental variables estimation and least squares estimation, greater diversity within the HLA system is shown to cause improvements in life expectancy before medicine. Without effective medicines and vaccines, individual resistance to infectious disease was dependent upon socioeconomic, geographic, and genetic traits. After controlling for necessary differences and using a valid instrument for genetic differences, heterozygosity within the HLA system is shown to provide an aggregate health advantage. The next section explores the sensitivity of this relationship through sample truncations and the inclusion of additional, potentially omitted, variables.

5.3 Robustness

5.3.1 Confirming 1960 as a Valid Measure of Premedicinal Health

The use of 1960 data is seen as a data trade-off, where the more prominent coverage and more accurate measures of 1960 are used in place of periods of time in which medicines and vaccines were still unknown. The complicating factor with the use of 1960 data is that the presence of medicine should dissipate any effect of inherent genetic resistance to infectious disease. To confirm this idea and to show that 1960 is an early enough period to capture the effects of inherent resistance, Table 6 explores the effects of HLA heterozygosity on life expectancy in more contemporary periods. If HLA heterozygosity does represent genetic resistance to infectious disease, the effect should be more pronounced in earlier periods. In the past 50 years, effective medicines and vaccines have more fully spread across the world. This greater presence of medicine should weaken the benefits of immune system differences in explaining variations in life expectancy.

Table 6 explores the effect of HLA heterozygosity on life expectancy from 1960 to 2010. If medicines are becoming more widely dispersed over time, then the effect of HLA heterozygosity, or inherent resistance, should begin to lessen. This idea is confirmed in Table 6. Column (1) reproduces our baseline regression with a sample representing all available life expectancy, GDP per capita, and years of schooling data for the periods considered. Both the OLS and 2SLS coefficient in column (1) are similar in magnitude to the estimates

found in column (5) of tables 4 and 5, respectively.

Columns (2)-(6) regress life expectancy in 1970-2010 (by decade) on HLA heterozygosity, respectively. As effective medicines become more widely distributed over time, the coefficient of HLA heterozygosity should move to zero. Comparing columns (2)-(6), the magnitude of the coefficient of HLA heterozygosity declines between 1960 and 1980 and then becomes insignificantly different than zero in 1990 and remains so until the present. This result is shown with both OLS and 2SLS estimates. The estimates of Table 6 support the use of 1960 as a valid measure for premedicinal health.

5.3.2 Sample Truncations

This subsection will firstly explore within continent estimations and secondly explore the effects of differing source populations, not geographic designations.

Table 7 restricts the sample to differing continents (or logical combinations) while using our baseline regression model, excluding continent indicators. Columns (1)-(3) restrict the sample to each Old World continent, respectively.[24] HLA heterozygosity has an insignificant effect on premedicinal life expectancy within Europe, Asia, and Africa, respectively.

The effect of HLA heterozygosity, however, is positive and statistically significant for column (4), which restricts the sample to countries within the Americas. The Americas contain large amounts of diversity between states due to historical migration patterns. In other words, the presence of indigenous populations and immigrants primarily from Europe and Africa allows for necessary variation between states to measure the effect of HLA heterozygosity; this is confirmed in the standard deviations of Table 1.

Columns (5) and (6) classify states based upon segregation before the Columbian Exchange and European colonization. The use of Old World states in column (5) suggest HLA heterozygosity is positively associated with premedicinal life expectancy; although, statistical significance dissipates for 2SLS estimates due to a lack of precision. Column (6) considers New World states, which simply includes five Oceanic states to the regression of column (4). Again, the greater amount of variation and the relatively large fractions of

[24] We consider Europe, Asia, and Africa Old World continents, with the Americas and Ocenia being New World continents.

indigenous American and Oceanic populations lead to more precise estimates of the effect of HLA heterozygosity.

One potential source of bias associated with our measure of HLA heterozygosity lies within the aggregation from ethnic groups to the country level. Given the limited number of ethnic groups from which HLA heterozygosity is constructed, our aggregated measure may simply be accounting for the fraction of a country's population that is derived from Eurasia. In other words, the relationship between HLA heterozygosity and life expectancy in 1960 may reflect some underlying role of Eurasia in promoting greater health outcomes. Aside from diversity in the HLA system, Eurasian populations may contain unseen cultural or additional genetic benefits. Therefore, it is worthwhile to explore the effect of HLA heterozygosity in countries of differing concentrations of Eurasian descent; this is considered in Table 8.

In order to account for the fraction of a country's population being from Eurasia, we use the continent associated with each ethnic group for which we have genetic data. Using ethnic classifications, we then observe the share of each country associated with an ethnic population from a given continent. If, for example, a country is composed equally of the French and Yoruba ethnic groups, the country is assigned a 50% fraction from Europe and 50% fraction from Africa. The fraction from Eurasia is then the fraction of a country's population from Europe plus the fraction from Asia.

Column (1) of Table 8 recreates the baseline estimation while including countries that are composed only of ethnic populations originating outside of Eurasian. Though the sample is small, the coefficient of interest is significant at the 5% level for both the OLS and 2SLS estimates, and the magnitude of the coefficients are similar to the baseline estimates. Column (2) restricts the sample to those countries that are composed partially of a population indigenous to Eurasia. The effect of HLA heterozygosity remains positive, similar in magnitude to the baseline estimates, and statistically significant at conventional levels. Column (3) restricts the sample to those countries solely composed of Eurasian populations. For this sample, the OLS estimate is similar in magnitude to previous estimates but is insignificantly different than zero. The 2SLS estimate, however, is significant at the 1%

level with a magnitude more than double the baseline estimate found in Table 5.

The estimations of Table 8 suggest that unseen benefits of Eurasian populations are not driving the relationship between HLA heterozygosity and premedicinal life expectancy. As a further control for the effect of Eurasian populations, we will include the fraction of the population derived from Eurasia into the baseline regression model. This is seen in the next sub-section.

5.3.3 Omitted Variables

Table 9 includes additional variables to the baseline estimation. The additional controls are intended to capture unseen influences on either HLA heterozygosity or life expectancy in 1960. In all regressions of Table 9, the coefficient of HLA heterozygosity is positive and statistically significant at conventional levels while the coefficient is slightly attenuated when the fraction of the population derived from Eurasia is included into the regression model.

Column (1) of Table 9 performs the baseline estimation for the adjusted sample that contains data for all potentially omitted variables. The sample adjustment doesn't significantly alter the coefficients of HLA heterozygosity.

As a further check against a potential bias from Eurasian populations in Table 8, column (2) of Table 9 includes the fraction of a countries population derived from Eurasia into the baseline regression model.[25] The fraction of a country's population derived from Eurasia has a positive and statistically significant effect on life expectancy in 1960, while the coefficient of HLA heterozygosity remains positive and similar in magnitude to the baseline estimate, although the magnitude is lessened in for both the OLS and 2SLS estimates. Given the results in column (2) of Table 9 and the estimates of Table 8, it is not likely that our measure of HLA heterozygosity is accounting for additional factors associated with Eurasian ancestry.

Column (3) includes a measure for the ease of transmission of malaria into the baseline estimation (Kiszewski et al. 2004). While this measure of malaria is negatively associated with premedicinal life expectancy, the effect is statistically insignificant; the negative point

[25] As an additional check, we've substituted the fraction of a country's population from Europe for the fraction from Eurasia.

estimate and insignificance occurs for both OLS and 2SLS estimation. The inclusion of this variable, however, does not weaken the effect of HLA heterozygosity. The coefficient of HLA heterozygosity is significant at the 1% level and similar in magnitude in to the baseline estimates.

Galor and Moav (2007) show a prolonged history of agriculture has a positive association with contemporary health outcomes. We argue, however, that the relationship between the Neolithic Revolution and contemporary health is due to prolonged exposure to crowd disease and resulting genetic adaptation. In other words, GM's weighted millennia of agricultural is a general measure of adaptation, whereas our measure is a direct result of adaptation. An ancestry adjusted measure for the millennia a country has practiced agriculture is controlled for in column (4). The coefficient of HLA heterozygosity is unaffected.

Following the theory of Spolaore and Wacziarg (2009), a greater level of genetic distance from the technological frontier is associated with a slower diffusion of technology; this slow diffusion of technology, in turn, should be associated with lower production, lower medical technologies, and therefore, lower life expectancy. In order to accurately ensure HLA heterozygosity, which is based on genetic differences, isn't picking up an omitted effect of technological diffusion, it is necessary to control for this measure of genetic distance. Column (5) includes the genetic distance from the United States into the baseline regression model. The inclusion of genetic distance from the United States does not result in any significant changes in either the OLS or 2SLS estimated effect of HLA heterozygosity. This result indicates that our measure of HLA heterozygosity is not being driven by general genetic differentiation; instead, variation within the HLA system is picking up a specific effect of the genome on aggregate health outcomes.

Column (6) includes a number of demographic controls, which includes the fraction of the population living in an urban area, the population density, and the fraction of the population between the ages of 0 and 14. All demographic variables are for 1960. Population density and urbanization are intended to control for the ease in transmission of the infectious crowd diseases under consideration, while the fraction of the population aged between 0 and 14 is used to account for young populations, which may be more vulnerable to the considered

diseases. The inclusion of demographic controls does not alter the sign, significance, or magnitude of the coefficient of interest.

All additional controls are included in column (7) of Table 9. The inclusion of all controls results in the magnitude of the coefficient of HLA heterozygosity slightly attenuating but remaining statistically significant at conventional levels; this is true for both the OLS and 2SLS estimate.[26] Holding constant other potentially relevant factors does not alter the effect of HLA heterozygosity.

6 Conclusion

The Neolithic Revolution radically changed the environment of early humans. As a result of the new environment, and the large populations that resulted, pathogens previously constrained to animal populations came to infect the readily available human hosts. These pathogens selected for variation within the immune system. Those societies and peoples that came into contact with these pathogens at an earlier date, as well as allowing for a sustained presence through large populations, have had a greater selection for particular traits that provide resistance to the resulting diseases. This has resulted in a contemporary variation in disease resistance across countries, which corresponds to a variation in aggregate health measures.

Galor and Moav (2007) show that an earlier agricultural transition has given a head start in adaptation to the environmental shift. This head start, in turn, is associated with contemporary variations in aggregate health outcomes. We see the current work as an intermediary to that of GM, where we attempt to more narrowly define the adaptation resulting from the Neolithic Revolution. Towards this end, we explore genetic variation within the HLA system across countries, where we propose that HLA diversity corresponds to differences in historical disease environments and a measures adaptation to these disease environments.

[26]The attenuation of the OLS and 2SLS estimated coefficients is similar to that in column (2), which controls for the fraction of the population derived from Eurasia.

References

[1] D. Acemoglu and S. Johnson. Disease and development: the effect of life expectancy on economic growth, 2006.

[2] Daron Acemoglu, Simon Johnson, and James Robinson. The colonial origins of comparative development: An empirical investigation. *The American Economic Review*, 91(5):1369–1401, 2001.

[3] Alberto Alesina, Arnaud Devleeschauwer, William Easterly, Sergio Kurlat, and Romain Wacziarg. Fractionalization. *Journal of Economic Growth*, 8(2):155–194, 2003.

[4] R.M. Anderson and RM May. Infectious diseases of humans: dynamics and control. *New York*, page 757, 1991.

[5] M.A. Arroyo, W.B. Sateren, D. Serwadda, R.H. Gray, M.J. Wawer, N.K. Sewankambo, N. Kiwanuka, G. Kigozi, F. Wabwire-Mangen, M. Eller, et al. Higher hiv-1 incidence and genetic complexity along main roads in rakai district, uganda. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 43(4):440, 2006.

[6] Q. Ashraf and O. Galor. The'out of africa'hypothesis, human genetic diversity, and comparative economic development. *MIT Department of Economics Working Paper No. 11-08*, 2011.

[7] W. Bank. World development indicators. *World Bank*, 2010.

[8] Ian Barnes, Anna Duda, Oliver G Pybus, and Mark G Thomas. Ancient urbanization predicts genetic resistance to tuberculosis. *Evolution*, 65(3):842–8, Mar 2011.

[9] K.K. Bhatia, F.L. Black, T.A. Smith, M.L. Prasad, and G.N. Koki. Class i hla antigens in two long-separated populations: Melanesians and south amerinds. *American journal of physical anthropology*, 97(3):291–305, 1995.

[10] J.M. Blackwell, S.E. Jamieson, and D. Burgner. Hla and infectious diseases. *Clinical microbiology reviews*, 22(2):370, 2009.

[11] H. Bleakley. Disease and development: comments on acemoglu and johnson (2006). *Remarks delivered at the NBER Summer Institute on Economic Fluctuations and Growth, July*, 16:73–117, 2006.

[12] J.S. Brownstein, C.J. Wolfe, and K.D. Mandl. Empirical evidence for the effect of airline travel on inter-regional influenza spread in the united states. *PLoS medicine*, 3(10):e401, 2006.

[13] L.L. Cavalli-Sforza, P. Menozzi, and A. Piazza. *The history and geography of human genes*. Princeton Univ Pr, 1994.

[14] G.S. Cooke and A.V.S. Hill. Genetics of susceptibility to human infectious disease. *Nature Reviews Genetics*, 2(12):967–977, 2001.

[15] A.W. Crosby. *Ecological imperialism: the biological expansion of Europe, 900-1900*. Cambridge Univ Pr, 2004.

[16] D.M. Cutler, A.S. Deaton, and A. Lleras-Muney. The determinants of mortality, 2006.

[17] R. Dawkins. *The extended phenotype: the gene as the unit of selection*. 1982.

[18] R. Dawkins. *The Selfish Gene: –with a new Introduction by the Author*. Oxford University Press, USA, 2006.

[19] P.I.W. de Bakker, G. McVean, P.C. Sabeti, M.M. Miretti, T. Green, J. Marchini, X. Ke, A.J. Monsuur, P. Whittaker, M. Delgado, et al. A high resolution hla and snp haplotype map for disease association studies in the extended human mhc. *Nature genetics*, 38(10):1166, 2006.

[20] J. Diamond. *Guns, Germs, and Steel: The Fates of Human Societies*. W.W. Norton, 1998.

[21] A.P. Dobson and E.R. Carper. Infectious diseases and human population history. *Bioscience*, pages 115–126, 1996.

[22] O. Galor and O. Moav. The neolithic origins of contemporary variation in life expectancy. *Brown University Department of Economics Working Paper*, 14:2007, 2007.

[23] L. Guiso, P. Sapienza, and L. Zingales. Cultural biases in economic exchange, 2004.

[24] L. Guiso, P. Sapienza, and L. Zingales. Does local financial development matter?*. *Quarterly journal of Economics*, 119(3):929–969, 2004.

[25] M. Harrison. *Disease and the modern world: 1500 to the present day*. Polity, 2004.

[26] D.L. Hartl and A.G. Clark. *Principles of population genetics*, volume 116. Sinauer associates Sunderland, MA, 4 edition, 2006.

[27] J. Hawks, E.T. Wang, G.M. Cochran, H.C. Harpending, and R.K. Moyzis. Recent acceleration of human adaptive evolution. *Proceedings of the National Academy of Sciences*, 104(52):20753, 2007.

[28] D.A. Hibbs and O. Olsson. Geography, biogeography, and why some countries are rich and others are poor. *Proceedings of the National Academy of Sciences of the United States of America*, 101(10):3715, 2004.

[29] A.V.S. Hill. The immunogenetics of human infectious diseases. *Annual review of immunology*, 16(1):593–617, 1998.

[30] R. Horton, L. Wilming, V. Rand, R.C. Lovering, E.A. Bruford, V.K. Khodiyar, M.J. Lush, S. Povey, C.C. Talbot, M.W. Wright, et al. Gene map of the extended human mhc. *Nature Reviews Genetics*, 5(12):889–899, 2004.

[31] A.L. Hughes and M. Yeager. Natural selection at major histocompatibility complex loci of vertebrates. *Annual review of genetics*, 32(1):415–435, 1998.

[32] M.C. Inhorn and P.J. Brown. The anthropology of infectious disease. *Annual review of Anthropology*, 19:89–117, 1990.

[33] K.J.M. Jeffery and C.R.M. Bangham. Do infectious diseases drive mhc diversity? *Microbes and Infection*, 2(11):1335–1341, 2000.

[34] J. Klein et al. *Natural history of the major histocompatibility complex.* Wiley New York, 1986.

[35] C. McEvedy and R. Jones. *Atlas of world population history.* Penguin Books Ltd, Harmondsworth, Middlesex, England., 1978.

[36] W.H. McNeill. *Plagues and peoples.* Anchor, 1976.

[37] D. Meyer, R.M. Single, S.J. Mack, H.A. Erlich, and G. Thomson. Signatures of demographic history and natural selection in the human major histocompatibility complex loci. *Genetics*, 173(4):2121, 2006.

[38] S. Michalopoulos. The origins of ethnolinguistic diversity: Theory and evidence. *Department of Economics, Tufts University*, 2008.

[39] N. Nunn. The long-term effects of africa's slave trades*. *Quarterly Journal of Economics*, 123(1):139–176, 2008.

[40] SB Piertney and MK Oliver. The evolutionary ecology of the major histocompatibility complex. *Heredity*, 96(1):7–21, 2005.

[41] F. Prugnolle, A. Manica, and F. Balloux. Geography predicts neutral genetic diversity of human populations. *Current Biology*, 15(5):R159–R160, 2005.

[42] F. Prugnolle, A. Manica, M. Charpentier, J.F. Guégan, V. Guernier, and F. Balloux. Pathogen-driven selection and worldwide hla class i diversity. *Current Biology*, 15(11):1022–1027, 2005.

[43] L. Putterman. Agriculture, diffusion and development: Ripple effects of the neolithic revolution. *Economica*, 75(300):729–748, 2008.

[44] L. Putterman and D.N. Weil. Post-1500 population flows and the long-run determinants of economic growth and inequality*. *Quarterly Journal of Economics*, 125(4):1627–1682, 2010.

[45] H. Rajeevan, M.V. Osier, K.H. Cheung, H. Deng, L. Druskin, R. Heinzen, J.R. Kidd, S. Stein, A.J. Pakstis, N.P. Tosches, et al. Alfred: the alelle frequency database. update. *Nucleic Acids Research*, 31(1):270, 2003.

[46] S. Ramachandran, O. Deshpande, C.C. Roseman, N.A. Rosenberg, M.W. Feldman, and L.L. Cavalli-Sforza. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in africa. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44):15942, 2005.

[47] P C Sabeti, S F Schaffner, B Fry, J Lohmueller, P Varilly, O Shamovsky, A Palma, T S Mikkelsen, D Altshuler, and E S Lander. Positive natural selection in the human lineage. *Science*, 312(5780):1614–20, Jun 2006.

[48] T. Shiina, H. Inoko, and JK Kulski. An update of the hla genomic region, locus information and disease associations: 2004. *Tissue Antigens*, 64(6):631–649, 2004.

[49] K. Shillington. *History of Africa*. Bedford Books, 1989.

[50] RW Slade and HI McCallum. Overdominant vs. frequency-dependent selection at mhc loci. *Genetics*, 132(3):861, 1992.

[51] E. Spolaore and R. Wacziarg. The diffusion of development*. *Quarterly Journal of Economics*, 124(2):469–529, 2009.

[52] J.A. Traherne, R. Horton, A.N. Roberts, M.M. Miretti, M.E. Hurles, C.A. Stewart, J.L. Ashurst, A.M. Atrazhev, P. Coggill, S. Palmer, et al. Genetic analysis of completely sequenced disease-associated mhc haplotypes identifies shuffling of segments in recent human history. *PLoS genetics*, 2(1):e9, 2006.

[53] M.E. Wilson. Travel and the emergence of infectious diseases. *Emerging infectious diseases*, 1(2):39, 1995.

[54] N.D. Wolfe, C.P. Dunavan, and J. Diamond. Origins of major human infectious diseases. *NATURE-LONDON-*, 447(7142):279, 2007.

[55] S. WRIGHT. *Evolution and the genetics of population (vol 4) variability within and among natural populations, paper*. 1984.

Appendix: Variable Definitions and Sources (Alphabetical Order)

Absolute Latitude

The absolute value of a country's representative latitude. Representative latitude is given by the centroid latitude of a country from *The World Factbook* (2011).

Ethnic Fractionalization

An index from Alesina et al. (2003), which represents the probability of two randomly selected individuals belonging to different ethnicities.

Fraction of Population Derived from Eurasia

Using the ethnicities for which we have data on HLA heterozygosity, ethnic groups are ascribed to the continent in which they are found. These data are aggregated to the country level from ethnic compositions in Alesina et al. (2003) in the same manner as HLA heterozygosity. This gives a country-level population compositions for each of our six continents: Africa, Europe, Asia, North America, South America, and Oceania. The fraction from Eurasia is the fraction of a country's population from Europe plus the fraction from Asia.

Fraction of Young Population in 1960

The fraction of a country's population that is aged 0-14 years in 1960. This variable is included in the vector of demographic controls and comes from the *World Development Indicators* (World Bank 2012).

Fraction within Tropics

The fraction of a country with a Kppen-Geiger tropical climate. These data come from Nunn and Puga (2012).

GDP per capita for 1960 and 1970

Maddison estimates for PPP converted GDP per capita in constant 2007 constant dollars. Found in Avakov (2010).

GDP per capita for 1980-2010

PPP converted GDP per capita in constant 2005 constant dollars. Data come from the World Development Indicators (World Bank 2012).

Genetic Distance from the U.S.

Genetic distance is a measure of genetic diversity between societies. This measure is calculated with the fixation index, or F_{ST}, from population genetics and measures the variation in gene frequencies across differing groups. F_{ST} scores are given for 42 indigenous populations; the data come from Cavalli-Sforza et al. (1994). The genetic distance measures are then aggregated to the country level by Spolaore and Wacziarg (2009), from which genetic distance to the USA is found for 206 countries. The USA is chosen as the technology frontier in 1500 CE. Genetic distance from this frontier is intended to convey difficulty in the diffusion of technology.

HLA Heterozygosity

This measure is discussed in detail in Section 3. In short, ethnic variation, which is measured by expected heterozygosity, is calculated for 156 SNP's found within genes associated with the HLA system (Kidd et al. 2003). These ethnic data are then aggregated to the country level with ethnic compositions in Alesina et al. (2003). This measure captures variation within the the part of the genome associated with the recognition and disposal of foreign pathogens.

Life Expectancy in 1940

Life expectancy in 1940 represents life expectancy at birth. These data come from historic UN and League of Nations reports by way of Acemoglu and Johnson (2007).

Life Expectancy in 1960-2010

Life expectancy represents life expectancy at birth. These data come from the *World Development Indicators* (World Bank 2012).

Malaria Ecology Index

The malaria ecology index takes into account differences in the environment and mosquito vectors that contribute to the spread of malaria. These data come from Kiszewski et al. (2004).

Migratory Distance from East Africa

This measure is constructed from ethnic migratory distances from East Africa found in Ashraf and Galor (2012). Ethnic groups from Ashraf and Galor are matched by language to groups for which we have HLA heterozygosity measures. These data are then aggregated to the country level in an identical manner to HLA heterozygosity. In effect, this creates an ancestry adjusted measure for historic migratory distance.

Migratory distance itself is measured by the great circle distance between an ethnic groups current position and the closest way point specified in Figure 2. This distance is then added to the distance from East Africa along a proposed migratory path out of Africa (Ashraf and Galor 2012; Prugnolle et al. 2005a; Ramachandran et al. 2005).

Millennia of Agriculture

The millennia since the majority of a country's population adopted agriculture for subsistence. These data are from Putterman (2008). Millennia of agriculture data are adjusted for migration movements between 1500-1960 through the use of a modified version of the Putterman and Weil (2010) migration matrix (Chanda et al. 2012).

Number of Potential Domesticate Animals

The number of prehistoric, native animals that were a potential source of domestication within a country. These data are from Hibbs and Olsson (2004). The number of potential domesticate animals is then adjusted for migration movements between 1500-1960 with the use of a modified version of the Putterman and Weil (2010) migration matrix (Chanda et al. 2012).

Population Density in 1 CE

Population data for 1, 1000, and 1500 CE come from McEvedy and Jones (1978). Land area for each country is based on contemporary borders and is from the *World Development Indicators*. These data are adopted from Ashraf and Galor (2011). Historic population density data are adjusted for migration movements between 1500-1960 with the use of a modified version of the Putterman and Weil (2010) migration matrix (Chanda et al. 2012).

Population Density in 1960

The number of persons per square kilometer for a country in 1960. This variable is included in the vector of demographic controls and comes from the *World Development Indicators* (World Bank 2012).

Urbanization in 1960

The fraction of the population living within an urban area in 1960. This variable is included in the vector of demographic controls and comes from the *World Development Indicators* (World Bank 2012).

Years of Schooling 1960-2010

Years of schooling measures the average years of schooling for a country's 15 and over population. These data come from Barro and Lee (2010).

7 Tables

Table1: Summary Statistics for Baseline Variables

Variable:	N	Mean	Std. Dev.	Min	Max
HLA Heterozygosity	124	0.3183	0.0218	0.2347	0.3529
Continent:					
Europe	31	0.3351	0.0101	0.3184	0.3529
Asia	30	0.3143	0.0149	0.2711	0.3298
Africa	31	0.3187	0.0158	0.2844	0.3352
Americas	27	0.3092	0.0242	0.2588	0.3503
Oceania	5	0.2848	0.048	0.2347	0.3439
Life Expectancy in 1960	124	55.1291	11.7686	31.1261	73.5498
GDP per capita in 1960	124	5246.781	6573.047	425.0004	48635.06
Years of Schooling in 1960	124	3.6265	2.5154	0.1342	10.0201
Ethnic Fractionalization	124	43.306	25.0736	1.1998	94.0175
Absolute Latitude	124	27.0345	16.9065	1	64
Fraction within Tropics	124	38.8036	44.8792	0	100
Out of Africa Migratory Dist.	124	6991.603	3405.501	2883.02	18608.56

Table 2: Explaining HLA Heterozygosity

Dependent Variable: ln HLA Heterozygosity				
	Out of Africa	Historical Environment		
	(1)	(2)	(3)	(4)
ln Migratory Dist. from East Africa	2.1383***			
	(0.2957)			
ln Mig. Dist. Sqr.	-0.1264***			
	(0.0169)			
ln Population Density in 1 CE		0.0207***		-0.0558***
(Ancestry Adjusted)		(0.0068)		(0.0099)
No. of Potential Domesticate Animals			0.0116***	0.0114***
(Ancestry Adjusted)			(0.0022)	(0.0030)
Pop. Density × Animals				0.0089***
				(0.0015)
Dist. for Max HLA Het.	4,532 km	–	–	–
N	124	76	76	76
R Sqr.	0.5354	0.1008	0.2615	0.4446

Summary: This table displays the relationship of factors associated with HLA heterozygosity. The migratory distance out of East Africa has been shown to be associated with genetic variation. Given the presence of disease in Eurasia, we expect a non-linear relationship between HLA heterozygosity and migratory distance; this is displayed in column (1). In addition to migratory distance, historical biogeographic differences resulted in differential exposure, and historical populations allowed for the sustainability of infectious crowd diseases. Columns (2) and (3) give the relationship with historic population density and the number of potentially domesticate animals, respectively. Column (4) gives the interaction between historic population and domesticate animals.

Notes: (i) HLA heterozygosity is a measure for genetic variation with the HLA system, which is responsible for the recognition of foreign pathogens. (ii) The migratory distance from East Africa associated with the greatest measure of HLA heterozygosity is 4,532 km, which is similar to the migratory distance of Grenada (country-level migratory distance measures are population weighted). (iii) OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses.

Table 3: Health Outcomes before the Epidemiological Transition

Dependent Variable:	Predicted Mortality in 1940			Life Expectancy in 1940		
	(1) OLS	(2) OLS	(3) 2SLS	(4) OLS	(5) OLS	(6) 2SLS
ln HLA Heterozygosity	-5.4713***	-3.3637***	-2.9433***	2.3648***	1.2733***	1.3817***
	(0.7963)	(0.8687)	(0.9160)	(0.3392)	(0.3976)	(0.4107)
ln GDP per Capita in 1940		-0.4490***	-0.4712***		0.2325***	0.2268***
		(0.1106)	(0.1058)		(0.0413)	(0.0434)
N	47	47	47	47	47	47
R^2	0.4056	0.5883	–	0.4587	0.7554	–
First Stage F-Stat.:	–	–	184.1147	–	–	184.1147
Overid. p-value:	–	–	0.0042	–	–	0.6877

Summary: This table displays the relationship between HLA heterozygosity and Acemoglu and Johnson's (2007) measure of health before the epidemiological transition. Columns (1)-(3) show the relationship between HLA heterozygosity and predicted mortality from a number of infectious diseases, while columns (4)-(6) show the relationship with life expectancy in 1940.

Notes: (i) HLA heterozygosity is a measure for genetic variation with the HLA system, which is responsible for the recognition of foreign pathogens. (ii) The 2SLS regressions in columns (3) and (6) use migratory distance from East Africa and its square as instruments for HLA heterozygosity. (iii) OLS coefficients are reported in columns (1), (2), (4), and (5), while 2SLS coefficients are found in columns (3) and (6). *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. (iv) The first stage, or Kleibergen-Paap, F-statistics satisfy the Stock-Yogo criteria for strong instruments; the p-value for the overidentifying restrictions test corresponds to Hansen's J Statistic.

Table 4: Baseline Estimation–OLS

Dependent Variable: ln Life Expectancy in 1960					
	(1)	(2)	(3)	(4)	(5)
ln HLA Heterozygosity	1.2508***	1.0380***	0.4980***	0.9943***	0.4951***
	(0.2434)	(0.1764)	(0.1435)	(0.1984)	(0.1487)
ln GDP per Capita in 1960			0.0499***		0.0497***
			(0.0127)		(0.0134)
ln Years of Schooling in 1960			0.1087***		0.1088***
			(0.0161)		(0.0161)
ln Ethnic Fractionalization			-0.0138		-0.0136
			(0.0093)		(0.0099)
ln Abs. Latitude				0.0169	0.0010
				(0.0286)	(0.0199)
Fraction within Tropics				0.0000	-0.0000
				(0.0007)	(0.0004)
Continent Dummies	N	Y	Y	Y	Y
N	124	124	124	124	124
R^2	0.1665	0.7155	0.8751	0.7180	0.8752

Summary: The baseline, OLS estimates are given in Table 4. Our main hypothesis is that HLA variation led to differential health outcomes before the widespread distribution of medicine, for which life expectancy in 1960 is seen as a proxy. Column (1) gives the bivariate relationship; column (2) includes continent fixed effects; column (3) includes the vector of socioeconomic controls; column (4) includes geographic controls; and column(5) gives the full baseline model.

Notes: (i) HLA heterozygosity is a measure for genetic variation with the HLA system, which is responsible for the recognition of foreign pathogens. (ii) Continent fixed effects include indicators for Europe, Africa, Asia, North America, and South America. (iii) GDP per capita in 1960 is found using Maddison's estimates. (iv) OLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses.

Table 5: Baseline Estimation–2SLS

Dependent Variable: ln Life Expectancy in 1960					
	(1)	(2)	(3)	(4)	(5)
ln HLA Heterozygosity	0.3685	0.8730***	0.6476***	0.8374***	0.6689***
	(0.2773)	(0.2843)	(0.1836)	(0.3082)	(0.1843)
ln GDP per Capita in 1960			0.0457***		0.0465***
			(0.0129)		(0.0134)
ln Years of Schooling in 1960			0.1079***		0.1072***
			(0.0161)		(0.0162)
ln Ethnic Fractionalization			-0.0114		-0.0116
			(0.0093)		(0.0101)
ln Abs. Latitude				0.0147	0.0037
				(0.0284)	(0.0198)
Fraction within Tropics				-0.0001	0.0001
				(0.0007)	(0.0004)
N	124	124	124	124	124
First Stage F-Stat:	53.5412	51.7272	80.5578	66.9878	96.2758
Overid. p-value:	0.0000	0.4652	0.4121	0.3484	0.4301

Summary: The baseline, 2SLS estimates are given in Table 5. Given the natural relationship between migratory paths and genetic variation, we use the distance from East Africa to correct for measurement error for HLA heterozygosity. The columns in Table 5 mirror those in Table 4.

Notes: (i) HLA heterozygosity is a measure for genetic variation with the HLA system, which is responsible for the recognition of foreign pathogens. (ii) Continent fixed effects include indicators for Europe, Africa, Asia, North America, and South America. (iii) GDP per capita in 1960 is found using Maddison's estimates. (iv) 2SLS coefficients are reported in each column. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. (v) The first stage, or Kleibergen-Paap, F-statistics satisfy the Stock-Yogo criteria for strong instruments; the p-value for the overidentifying restrictions test corresponds to Hansen's J Statistic.

Table 6: Additional Years: The Effect of Medicine

Dependent Variable: ln Life Expectancy						
Year:	1960	1970	1980	1990	2000	2010
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: OLS Estimates			
ln HLA Heterozygosity	0.5955***	0.4533***	0.3423***	0.0867	0.0710	0.0667
	(0.1538)	(0.1485)	(0.1159)	(0.1305)	(0.1038)	(0.0950)
Baseline Controls	Y	Y	Y	Y	Y	Y
Continent Dummies	Y	Y	Y	Y	Y	Y
N	96	96	96	96	96	96
R^2	0.8819	0.8940	0.9048	0.8648	0.8474	0.8127
			Panel B: 2SLS Estimates			
ln HLA Heterozygosity	0.6819***	0.3567*	0.3156**	0.0028	-0.0644	-0.0254
	(0.1896)	(0.1943)	(0.1565)	(0.1529)	(0.1298)	(0.1130)
Baseline Controls	Y	Y	Y	Y	Y	Y
Continent Dummies	Y	Y	Y	Y	Y	Y
N	96	96	96	96	96	96
First Stage F-Stat:	104.0298	96.6801	98.8788	103.2751	102.7730	102.4484
Overid. p-value:	0.2270	0.6615	0.4682	0.1325	0.0837	0.0837

Summary: This table supports 1960 as a valid proxy for pre-medicina health. This is shown by the declining coefficient of HLA heterozygosity, which has an insignificant effect on life expectancy after 1980. The distribution of medicine negates the benefits of inherent resistance. More contemporary periods are associated with a greater prevalence and use of medicines that lead to the epidemiological transition; therefore, our measure of inherent resistance becomes insignificant.

Notes: (i) HLA heterozygosity is a measure for genetic variation with the HLA system, which is responsible for the recognition of foreign pathogens. (ii) The baseline controls include GDP per capita, years of schooling, ethnic fractionalization, absolute latitude, and the fraction of a country within the tropics. (iii) The measures for GDP per capita and years of schooling vary with the year of life expectancy. For GDP per capita in 1960 and 1970 Maddison's estimates are used, while for 1980-2010 GDP per capita is given by the World Bank. (iv) Continent fixed effects include indicators for Europe, Africa, Asia, North America, and South America. (v) OLS coefficients are reported for each column in Panel A, while 2SLS coefficients are reported in each column for Panel B. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. (vi) The first stage, or Kleibergen-Paap, F-statistics satisfy the Stock-Yogo criteria for strong instruments; the p-value for the overidentifying restrictions test corresponds to Hansen's J Statistic.

Table 7: Within Continent Effects

Dependent Variable: ln Life Expectancy in 1960						
Continent:	Europe	Asia	Africa	Americas	Old World	New World
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: OLS Estimates			
ln HLA Heterozygosity	-0.0166	-0.3557	0.6540	0.6443**	0.4923*	0.4606***
	(0.1755)	(0.4839)	(0.5511)	(0.2307)	(0.2808)	(0.1336)
Baseline Controls	Y	Y	Y	Y	Y	Y
Continent Dummies	N	N	N	N	N	N
N	31	30	31	27	92	32
R^2	0.5548	0.6527	0.5614	0.8258	0.8401	0.8603
			Panel B: 2SLS Estimates			
ln HLA Heterozygosity	-1.9538	-0.2330	0.4511	0.6700***	0.5068	0.5639***
	(1.8330)	(0.9301)	(0.9252)	(0.2252)	(0.4922)	(0.1892)
Baseline Controls	Y	Y	Y	Y	Y	Y
Continent Dummies	N	N	N	N	N	N
N	31	30	31	27	92	32
First Stage F-Stat:	1.6154	11.6557	8.4109	229.4280	20.0670	80.2285
Overid. p-value:	0.7977	0.2038	0.1355	0.5559	0.0400	0.3481

Summary: This table gives the within continent effects of HLA heterozygosity on life expectancy in 1960. Column (1) restricts the sample to European countries; column (2) restricts the sample to Asian countries; column (3) restricts the sample to African countries; column (4) considers countries with the Americas; and columns (5) and (6) consider Old World vs. New World effects.

Notes: (i) HLA heterozygosity is a measure for genetic variation with the HLA system, which is responsible for the recognition of foreign pathogens. (ii) The baseline controls include GDP per capita, years of schooling, ethnic fractionalization, absolute latitude, and the fraction of a country within the tropics. (iii) Continent fixed effects are excluded. (iv) OLS coefficients are reported for each column in Panel A, while 2SLS coefficients are reported in each column for Panel B. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. (v) The first stage, or Kleibergen-Paap, F-statistics satisfy the Stock-Yogo criteria for strong instruments; the p-value for the overidentifying restrictions test corresponds to Hansen's J Statistic.

Table 8: Truncation Based on Fraction of Population Derived from Eurasia

Dependent Variable: ln Life Expectancy in 1960			
	(1)	(2)	(3)
Fraction from Eurasia:	= 0%	∈ (0%, 100%)	= 100%
	Panel A: OLS Estimates		
ln HLA Heterozygosity	0.5604**	0.4130**	0.4757
	(0.2514)	(0.1622)	(0.3107)
Baseline Controls	Y	Y	Y
Continent Dummies	N	N	N
N	21	42	61
R^2	0.5303	0.8080	0.8148
	Panel B: 2SLS Estimates		
ln HLA Heterozygosity	0.6875**	0.4309**	1.6650***
	(0.2767)	(0.1984)	(0.6116)
Baseline Controls	Y	Y	Y
Continent Dummies	N	N	N
N	21	42	61
First Stage F-Stat:	225.7491	123.2666	16.8678
Overid. p-value	0.0722	0.1833	0.1579

Summary: This table explores the effect of HLA heterozygosity for differing Eurasian population frequencies. Eurasia contained advantages in initiating and sustaining infectious crowd diseases. This lead to the selection of variation within the HLA system. Eurasia may have contained other benefits, however. We check for this by truncating the sample based on Eurasian populations. Column (1) restricts the sample to countries with no fraction of their population being derived from Eurasia; column (2) restricts the sample to with some fraction of the population derived from Eurasia; and column (3) restricts the sample to countries for which the entire population is derived from Eurasia.

Notes: (i) HLA heterozygosity is a measure for genetic variation with the HLA system, which is responsible for the recognition of foreign pathogens. (ii) Countries with no fraction from Eurasia are mostly from Sub-Saharan Africa. Countries with some fraction are mostly within the Americas. Countries with complete Eurasian populations are mostly within Eurasia. (iii) The baseline controls include GDP per capita, years of schooling, ethnic fractionalization, absolute latitude, and the fraction of a country within the tropics. (iv) Continent fixed effects are excluded. (v) OLS coefficients are reported for each column in Panel A, while 2SLS coefficients are reported in each column for Panel B. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. (vi) The first stage, or Kleibergen-Paap, F-statistics satisfy the Stock-Yogo criteria for strong instruments; the p-value for the overidentifying restrictions test corresponds to Hansen's J Statistic.

Table 9: Omitted Variables

Dependent Variable: ln Life Expectancy in 1960							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: OLS Estimates			
ln HLA Heterozygosity	0.5309***	0.3668**	0.5157***	0.5013***	0.5174***	0.5110***	0.3777**
	(0.1514)	(0.1603)	(0.1554)	(0.1521)	(0.1553)	(0.1540)	(0.1638)
Additional Controls:							
Frac. of Pop. from Eurasia	N	Y	N	N	N	N	Y
Malaria Ecology Index	N	N	Y	N	N	N	Y
Millennia of Agriculture	N	N	N	Y	N	N	Y
Genetic Dist. from USA	N	N	N	N	Y	N	Y
Demographics	N	N	N	N	N	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y	Y
Continent Dummies	Y	Y	Y	Y	Y	Y	Y
N	109	109	109	109	109	109	109
R^2	0.8800	0.8889	0.8818	0.8845	0.8803	0.8920	0.8970
				Panel B: 2SLS Estimates			
ln HLA Heterozygosity	0.6931***	0.5906***	0.7350***	0.6521***	0.6807***	0.6880***	0.6208***
	(0.1817)	(0.2022)	(0.1759)	(0.1847)	(0.1907)	(0.1816)	(0.2123)
Additional Controls:							
Frac. of Pop. from Eurasia	N	Y	N	N	N	N	Y
Malaria Ecology Index	N	N	Y	N	N	N	Y
Millennia of Agriculture	N	N	N	Y	N	N	Y
Genetic Dist. from USA	N	N	N	N	Y	N	Y
Demographics	N	N	N	N	N	Y	Y
Baseline Controls	Y	Y	Y	Y	Y	Y	Y
Continent Dummies	Y	Y	Y	Y	Y	Y	Y
N	109	109	109	109	109	109	109
First Stage F-Stat:	87.5366	116.1079	97.2103	83.7298	76.8776	82.6109	83.0158
Overid. p-value:	0.3095	0.3168	0.2617	0.2846	0.3156	0.1993	0.2326

Summary: This table conditions on potentially omitted variables. These include the fraction of a country's population originating in Eurasia, a measure for the transmission of malaria, an ancestry adjusted measure for the length of time a country has practiced agriculture, the genetic distance from the United States, and demographic controls, which include population density in 1960, urbanization in 1960, and the fraction of the population betwee 0 and 14 years of age.

Notes: (i) HLA heterozygosity is a measure for genetic variation with the HLA system, which is responsible for the recognition of foreign pathogens. (ii) The baseline controls include GDP per capita, years of schooling, ethnic fractionalization, absolute latitude, and the fraction of a country within the tropics. (iii) Continent fixed effects are excluded. (iv) OLS coefficients are reported for each column in Panel A, while 2SLS coefficients are reported in each column for Panel B. *, **, and *** represent significance at the 10, 5, and 1% significance level, respectively. Robust standard errors are in parentheses. (v) The first stage, or Kleibergen-Paap, F-statistics satisfy the Stock-Yogo criteria for strong instruments; the p-value for the overidentifying restrictions test corresponds to Hansen's J Statistic.

8 Figures

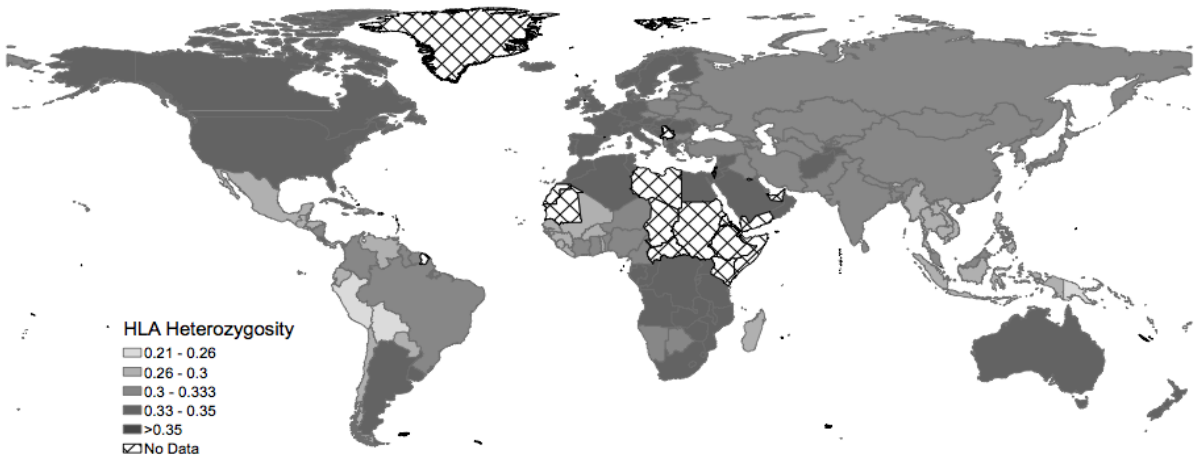

Figure 1
Cross-country HLA Heterozygosity

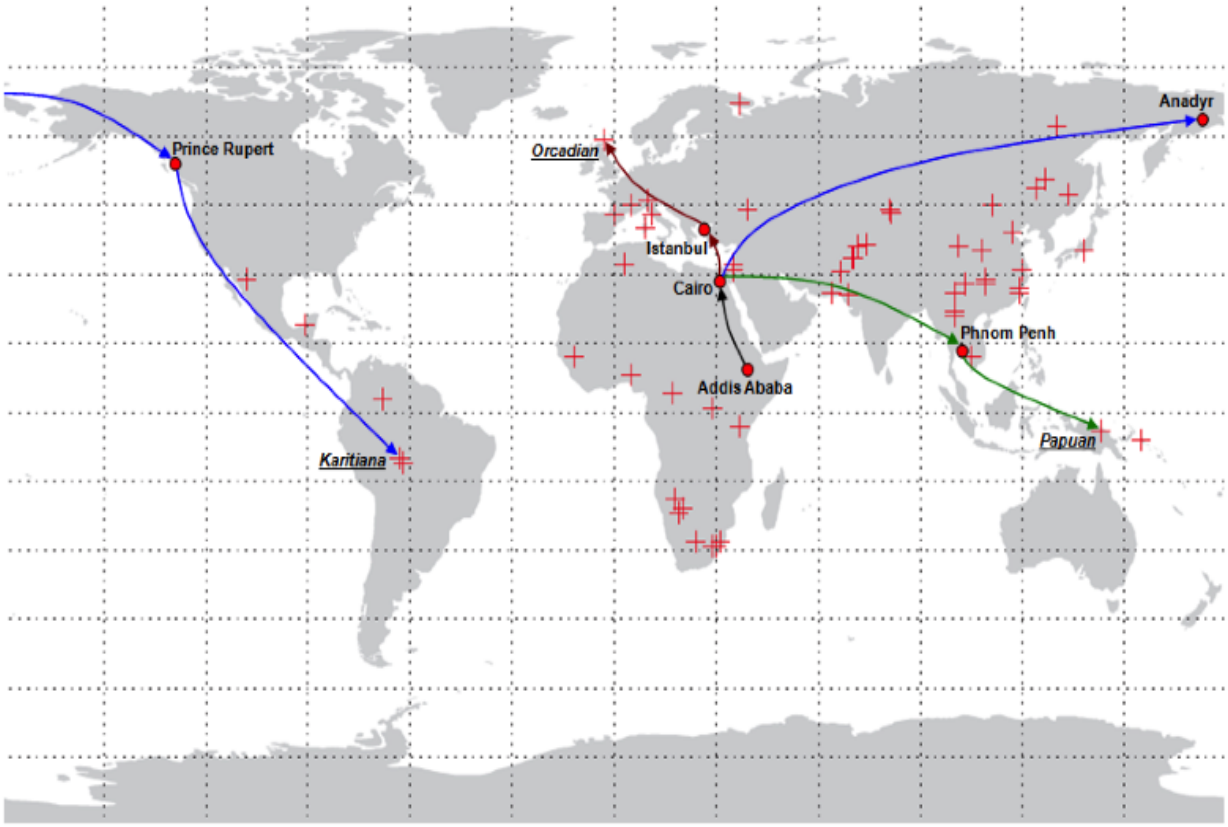

Figure 2
Migratory Paths from East Africa
(from Ashraf and Galor *forthcoming*)

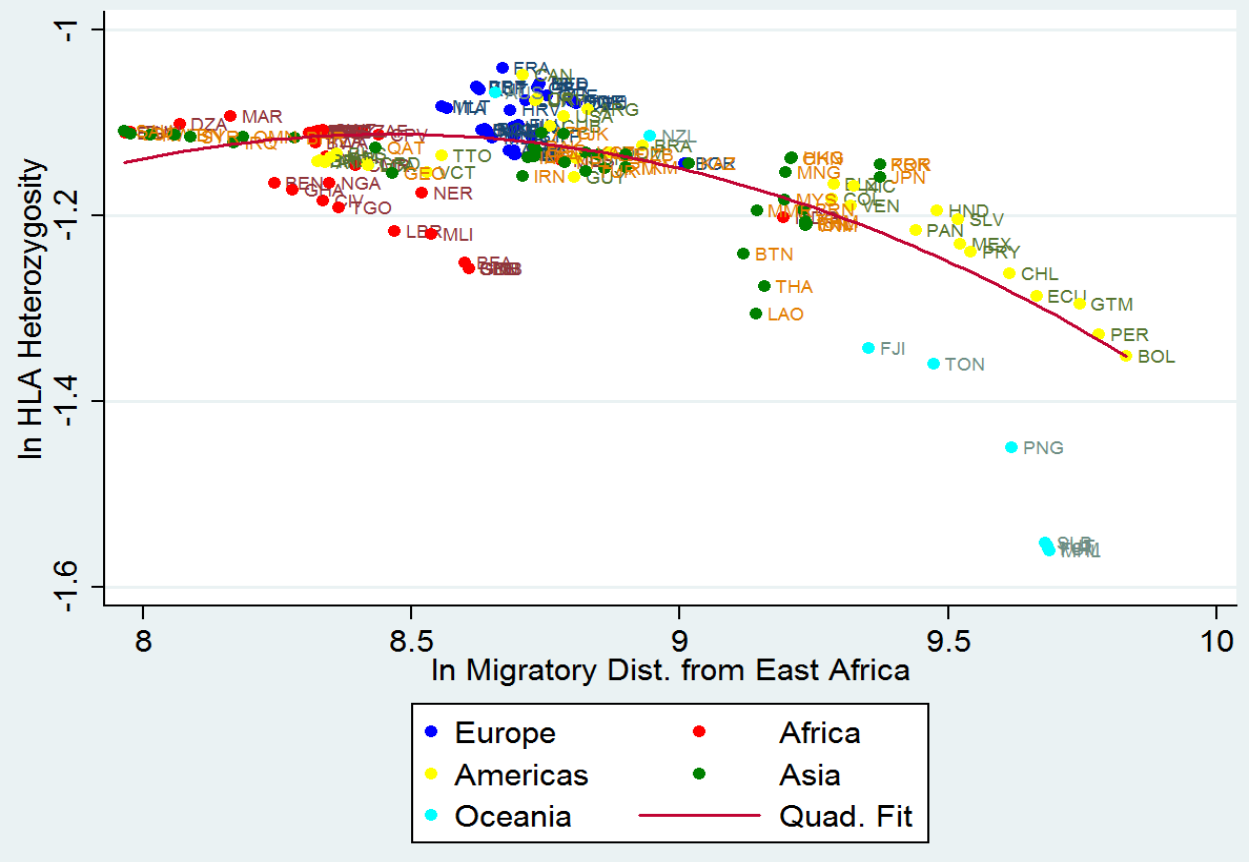

Figure 3
Relationship of "Out of Africa" Mig. Dist. and HLA Heterozygosity