ANCESTRYbyDNA 2.0 USER MANUAL

TABLE OF CONTENTS:

DNAPrint genomics, Inc.
900 Cocoanut Ave., Sarasota, FL 34236 tel 941-366-3400 fax 941-952-9770
www.ancestrybydna.com
email: admin@dnaprint.com.

What follows is the User Manual for the genomics test you have purchased. To fully understand your results, we recommend that you read this user manual from the beginning to the end. **To view your data files you will need the Adobe Acrobat Reader 5.0 Program available free of charge at http://www.adobe.com/products/acrobat/readstep2.html**.

How the test works:

To determine your ancestry, we have extracted DNA from your buccal sample and used the ***ANCESTRYbyDNA 2.0*** test to determine the sequence of your DNA at a large number of different positions. The buccal sample you returned to us contained thousands of cells, and each of your cells contains your DNA. Though we are all 99.9% identical at the level of our DNA sequence, there are certain regions of each chromosome that are different from person to person. These regions are called genetic markers or Single Nucleotide Polymorphisms (SNPs), and it so happens that some small fraction of these SNPs are also different amongst the world's continental population groups. These types of markers are best termed Ancestry Informative Markers (AIMs) and they constitute less than 5% of our genetic material, which is related to the very recent common origin of our species. To help you better understand why this is the case, we recommend the book: The Great Human Diasporas; The History of Diversity and Evolution, by Cavalli-Sforza. The positions in your DNA we have sequenced are those we have discovered that are different in this way, and they are spread across all of your chromosomes. That is why it is called a "genomics" test – it is a survey of all of your genetic material, which is known as your "genome". In other words, we have sequenced markers from chromosome 1, 2, 3 … 22, X and Y. Until ***ANCESTRYbyDNA 2.0***, genetics tests for genealogy or personal interest have been restricted to just one chromosome, like Y-chromosome or the mitochondrial DNA. As such, these other tests offer information that is very different from ***ANCESTRYbyDNA 2.0***. It is not that this information is incomplete or defective, its simply different information and it is useful for other purposes.

Your DNA was derived from your mother and father, and theirs was derived from their mothers and fathers, and theirs from their parents and so on. You have 23 pairs of chromosomes. Your maternal copy of chromosome 1 could have been passed through your mother from your maternal grandmother OR your maternal grandfather, but which one you received was randomly determined at conception (you could not have received both). Most of the time this copy chromosome 1 that you receive from your mother is actually a chimeric chromosome that includes parts from your grandfather and your grandmother. Recombination is the process by which these chimeric chromosomes are created and occurs at least once on each chromosome every time a new sperm or egg cell is made. As such, although blending does not occur at the level of the gene (the unit of trait expression) our chromosomes are mixed together and so our genomes contain segments of DNA from all of our ancestors. In contrast, the mitochondrial DNA or Y-chromosome test can only provide data on one single lineage of ancestors each

generation into the past. For example, 10 generations ago (year 1802 at 20 years per generation), a baby born today has 1024 ancestors. By measuring your ancestral proportions using a genomics method, we are actually measuring the average population affiliations of all of these 1024 ancestors. Since random processes (recombination and independent assortment) at inception determines the mixings and pairings you harbor, two offspring from a set of parents may have different sets of chromosome pairs, and therefore different ancestral proportions even though they were the product of the same male-female union. For example, the CSO of our company is a European male who married a Hispanic woman and had three children of mixed descent. Each of the children exhibits their own unique proportionality, which you can see by clicking on http://www.ancestrybydna.com/casestudy.pdf. If these two had an infinite number of children, the average would correspond to that proportionality exactly between the mother and father, but each child would deviate from this average by a unique and random amount.

Statement on Race.

Race is a defining issue of modern times in the US, Europe, and many other parts of the world. The impact of the European colonial period that started more than 500 years ago has set the tone for the interactions among diverse populations of the world. Colonization, genocide, Slavery, Legalized Segregation, Apartheid, Jim Crow Laws and Concentration Camps are but a few of the atrocities that are the history of our civilized world and every culture has its own list to be ashamed of. Given the enormity of these events, their long-term consequences will take generations to overcome. Modern conceptions of Race, Racism, and Racialization are some of the fallout of these events.

Part of our mission at DNAPrint is to work towards the abolition of these misconceptions and the social injustice that result from Racism and Racialization. In this light, we are dedicating considerable internal resources to education regarding the different perspectives (Sociocultural, Political, and Biological) on race and the meaning of populations in light of genomic science and biomedical research.

1) Race is not a biological concept. There is not enough genetic differentiation among human populations to consider them zoological races.
2) Race is a social construct. This means that these classifications (black, white, Hispanic, Jewish) are defined (and redefined) by the prevailing sociopolitical structure.
    a. Race is often a great amalgamation of many diverse populations and ethnicities.
    b. Race is often ascribed only to the minority populations.
    c. In the US, any minority population ancestry is dominant and the person is completely of the minority group (e.g. “the one-drop rule”).

3) Despite the veracity of points one and two above, since there is a correspondence amongst broad racial categories and populations, the conclusion that there are no average biological differences amongst any racially described groups may not be true.
4) Racism continues. "In some places, and for some people, overt racism has given way to implicit racialization and "Colorblind Racism" a term coined by Dr. Eduardo Bonilla-Silva (Stanford University)."
5) Race should not be used as a surrogate for population. Doing so may lead to over generalization and unfounded stereotypes. A population is the unit of evolution and refers to a group of persons who generally select mates from within the group.
6) Being respectful is the first important step in not having a racialized perspective.
    a. Each person is a human being first and foremost. It is disrespectful, at any level, on the street, in the lab, or in the clinic, to consider his or her population group first.
    b. Populations should be described (not defined) in precise language that members of the community would use.
7) We believe that the physical and cultural diversity of the world's peoples should be embraced as a valuable and even sacred resource. Indeed, the genomic variation both within and amongst populations is in many ways our Human Biodiversity and will provide important clues as to the origins, our physiological construction, and the possible futures of our fragile species.

Understanding BioGeographical Ancestry

The ANCESTRYbyDNA test provides a research grade estimation of a person's BioGeographical Ancestry. BioGeographical Ancestry (BGA) is a means of expressing the proportional ancestry of a person that is devoid of the ethnic labels and the dichotomous grouping of persons into racial categories. There are important uses of this in epidemiological and complex diseases mapping research and in forensic science. BGA estimates provide a description of a person in terms of ancestral proportions that are based on the evolutionary and geographical history of our species. Our recommended book, by one of the leaders in the field of Evolutionary Anthropology, Dr. Luca Cavalli-Sforza (Stanford University), details a broadly accepted model of human evolution. It is within this scientific framework of human origins that the BGA estimation can be understood as a description of a person's placement on a ***Multi-Dimensional Continuum of Ancestry*** ™.

What the results mean:

As you can see from the map provided, human beings migrated out of central, sub-Saharan Africa some 200,000 years ago to inhabit various regions of our globe. These migrants established founder groups that gave rise to present-day Europeans, Native Americans, Africans, and or East Asians/Pacific Islanders. A map of these human migration patterns can be found in

the map.gif file on your CDROM. If your heritage has been derived from more that one of these groups, that is, you are of mixed ancestry, the test results tell you what your mixture ratios are. If you do not have recent mixture, the test identifies which groups you are part of, and confirms that there is no evidence of recent admixture. You may be of admixed ancestry, or you may be of relatively unmixed ancestry. It so happens that many people from places such as Nigeria, Ireland and Japan are of relatively unmixed heritage (African, Indo-European and East Asian, respectively – see our website at www.ancestrybydna.com), but many people from places such the United States, South East Asia or Latin America are admixed. For example, Hispanics from Mexico or elsewhere in Central or South America were derived from the colonial mixture of Europeans, Native Americans, with some proportion of West African. Native Americans inhabited North and South America from Alaska to Patagonia. If your great grandfather was a great Aztec warrior, and of unmixed heritage, you will exhibit at least 12% Native American ancestry. If your great-great grandmother was a sanguine Chinese philosopher of unmixed heritage, you would be of at least 6% East Asian heritage. Many people from Puerto Rico are heavily admixed – showing significant Native American, African and Indo-European mixture. It is important to point out that, whether your heritage is a recently admixed or not, these results do not give you any information other than your ancestral proportions. For example, you should not use your results to make inferences about your predisposition to respond to a particular drug, or develop a particular disease. Additional tests that others and we are developing would be required for these purposes (they would measure specific sequences in your DNA that are relevant for each).

It is notable that some regions of the world have more complicated histories than other regions making the concept of ancestry more complicated and even tedious. For much of our history as a species, we were more mobile than today. The advent of agriculture, in at least four separate global regions about 10,000 years ago changed this for many people, but did not stop the process of migration. Indeed, the largest migrations in human history started only 500 years ago with the European colonial period, the trade in enslaved West Africans, and the colonization of the New World. However, prior to this time and for millennia people have moved about and particular regions of the world show traces of these migrations back and forth into and out of continental and sub-continental regions. Some examples of such regions are East Africa, North Africa, Central Asia, South Asia, and Insular Southeast Asia. Although these populations are distinct groups today with languages, cuisines and cultures that identify them as such, their genetic makeup reflects the long-term history of migrations from more than one region.

Your Ancestry Informative Marker Genotypes:

Your sequences are provided to you in the genotypes.doc file. Each of the sequences is comprised of two letters. There are two sequences because you received one chromosome from your mother and another from your father. DNA is comprised of nucleotide bases and there are four different types: Guanine (G), Adenine (A), Thymidine (T) and Cytosine (C). Most DNA

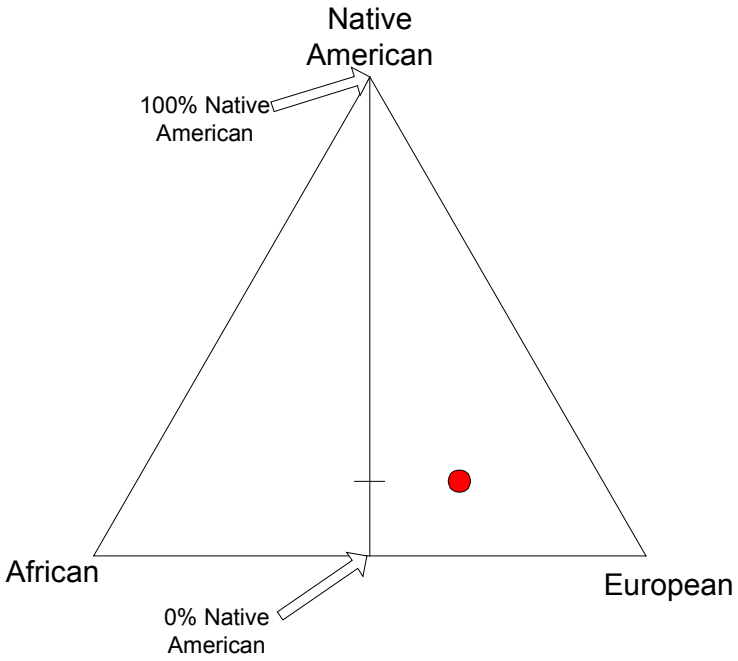
sites do not differ amongst individuals, but the sites we measure are variable in sequence from person to person, as well as from one population to the next. It is these positions that we measure order to estimate your ancestry. Each site has only two possible letters – they are called bi-allelic sites for this reason. Whereas one person may have two copies of a "G" at one site (represented as "G/G"), another may have a G and a C at this same site (represented as "G/C"). The possible letters for one site may be G or C, but for another they may be C or T, or A or G. Your string of letter pairs across all sites measured are most likely quite unique to you, and in a way they could represent a sort of a genetic bar code for your identification. However, it is the information they give us about your ancestral heritage that is of value here.

It so happens that these letters change slowly over time. When our ancestors migrated to establish the founder groups that gave rise to today's continental population groups, the nature of their sequences at these sites gradually drifted towards one or the other letter. By measuring the frequency of each letter in each of the descendent ancestral groups of these splinter groups, we can construct a sort of genetic map distinguishing each. We have sequenced your DNA at these sites and made an estimate of the proportional extent to which you share identity with each of four major population groups, sub-Saharan African, Indo-European, East Asian, and Native American.

Your Triangle Plots:

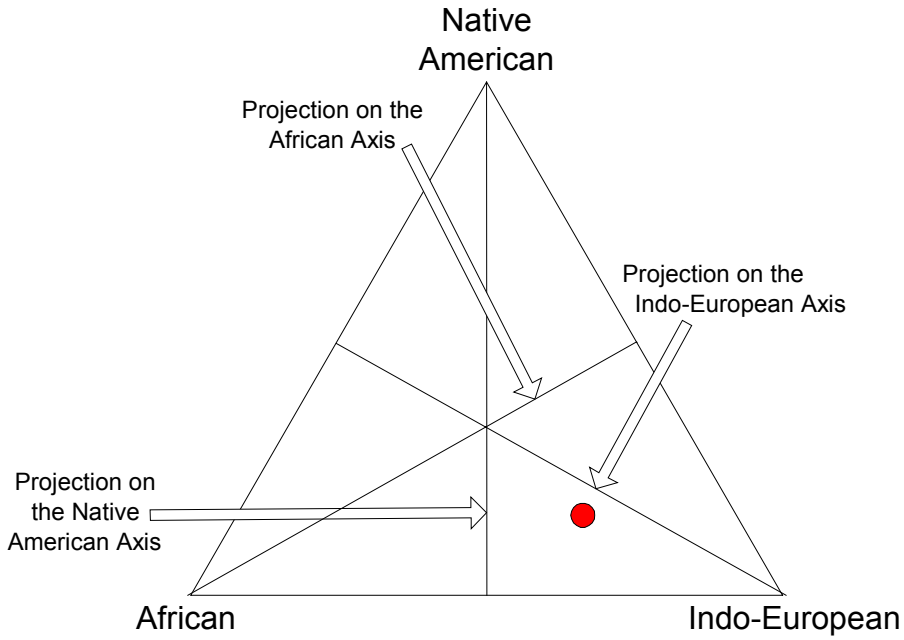
If your proportions are of three or fewer ancestries, we have provided you with a graphical "shapshot" of your results called a triangle plot. This plot is provided if your ancestry is derived from three or fewer groups (otherwise, it is impossible to represent your data in 2-dimensional space). The point on this plot is called a Maximum Likelihood Estimate (MLE), and only one is present on the plot. This MLE represents the best estimate of your ancestral proportions. To read a triangle plot, you drop a perpendicular line from the vertex (triangle point) of each triangle to the triangle edge below it.

DNAPrint genomics, Inc.
900 Cocoanut Ave., Sarasota, FL 34236 tel 941-366-3400 fax 941-952-9770
www.ancestrybydna.com
email: admin@dnaprint.com.



In the example above, the red dot is the MLE. We have dropped a line from the Native American vertex to the line below, and this particular line serves as a scale for the percentage of Native American ancestry – from 0% at the base to 100% at the vertex (or tip). Projecting the circle on this line is like holding a flashlight to the right of the triangle at the same level as the circle and observing the shadow the circle makes on the line. Where this shadow falls tells you the percentage of Native American ancestry. In this example, the individual is about 15% Native American (indicated by the hash mark).

DNAPrint genomics, Inc.
900 Cocoanut Ave., Sarasota, FL 34236 tel 941-366-3400 fax 941-952-9770
www.ancestrybydna.com
email: admin@dnaprint.com.



In the example above, we have created the scale line for each of the other two vertices. As with the first figure, for each line the point or vertex represents 100%, and the base near the line represents 0%. The arrows show where the circle projects on to each of these lines. This is sort of tricky to visualize, but if you held the flashlight in front of the point at a right angle to any one of the three 0%-100% scales you get the projection of the point on that scale, which gives you the percent value for that Ancestral group. In addition to being able to see that the person is of 15% Native American ancestry in the plot above, we can see that the person is of 60% Indo-European Ancestry and 25% African ancestry as well. You will notice that the three percentages must add up to 100%.

The MLE we have plotted is a statistical estimate, and all statistical estimates have confidence intervals. The MLE tells you what the best estimates of the proportions are, but in reality, there is a small chance that they are something else. The confidence intervals tell you what other values are also likely. In order to know your proportions with 100% confidence, we would have to perform the test for each region of the variable genome, which would make the test very expensive. Since we have not, your results are statistical estimates. We calculate and plot for you all of the estimates that are "2 fold, 5 fold and 10 fold less likely" than the MLE. The way it works is that the MLE is the most likely estimate. Any point within the first contour is up to "2 fold less likely" than the MLE, the farther from this point, the closer to it is to "2 fold less likely". Any point outside the first and within the second is from "2-5 fold less likely", again,

the farther from the MLE, the closer to "5-fold". Same with the last contour - any point within the last contour, but outside the second contour is from "5-10 fold less likely", the farther from the MLE, the closer to "10-fold". Any point outside the last contour is AT LEAST "10 fold less likely". Virtually every point outside the last contour is more than "10 fold less likely", the farther from the MLE, the greater the increase over "10 fold". The greater the number of DNA positions we read, the closer these contour lines come to the MLE point. On the triangle plot, the likelihood (probability) that your true value is represented by a different point, than the MLE decreases as you approach the red dot, where the probability is at its maximum (hence, it is called the Maximum Likelihood Estimate or MLE). We could perform the test so that the contour lines are very close to the MLE, however this would require us to sequence a much larger collection of markers. To keep the test affordable, we limit the survey a reasonable number of markers that are sufficient for you to know with good confidence what your proportions are. The yellow circle (10 fold contour) is also referred to as the "one-log interval" and is taken generally taken as a scientific level of confidence.

Your ancestry can only be plotted inside a triangle if you are of only one ancestry (i.e. unmixed) or your mix is composed of 2 or 3 ancestries. If your mix is of 4 ancestries, we cannot visualize your plot on a 2-dimensional sheet of paper (you are quite a unique blend of ancestries!), but you could make an *estimate* MLE in the triangle if most of your mix is due to 3 or fewer ancestries and the others combine to explain only a few percent.

Your Data Table and the results.doc file:

The triangles are tools for you to visualize your data. The actual data is provided to you in the form of a table such as this:

SAMPLE-ID = XXXXXXXX

ESTIMATE	ANCESTRY
90%	Indo-European
10%	Native American
0%	African
0%	East Asian

The percentage of Indo-European (EUR), African (AFR), East Asian (EAS), and Native American (NAM is shown. In the example above, the person is 90% Indo-European, 10%

Native American, and the MLE is plotted at the position of the triangle plot indicated by these proportions.

The ancestral proportions we report to you are a function of probability, and the MLE is the best estimate. Going back to the triangle, the further from the point estimate on the triangle, the less likely that point represents your true values. The farther from the red dot (the MLE) in any direction, the lower the probability. Therefore, your plot is an estimate rather than a precise fix, but it is the most likely estimate based on the data produced from many regions of your genome.

In other words:

a) The MLE is the most likely value. If you were to use one proportion to describe your heritage, it would be the MLE value and it is this value that appears on your certificate.
b) If this MLE value happens to be wrong, which is a possibility, it is ten times more likely that that your value is within the yellow circle than outside the yellow circle, or equivalently.

Why is it that it is not possible to determine what your proportions are *exactly*? When drawing conclusions from DNA one must use statistics. To do this without using statistics, we would have to go back in time 200,000 years and keep track of every one of your ancestors since the origin of our species in Africa. Clearly, this is impossible. Since you inherited your DNA from your ancestors, your ancestral proportions are written in your DNA, but this information must be statistically inferred from the DNA sequence. If we measured 1,000 or 10,000 genetic sites in your DNA as opposed to the 200 or so we measure today, your confidence intervals would be smaller but the MLE would probably be the same (in the triangle plot, the confidence ring would shrink around the MLE point). The costs of measuring more genetic sites can quickly become out of hand for most customers, and as with most things, there is a trade-off between expense and accuracy. It is sort of like a microscope. What you get from the lens is a representation of the object you are looking at. To truly know the details of the object, one would have to shrink themselves and get inside of the object in order to map all of its details.

Thus, in conclusion, the result of our test is your MLE. Though the MLE is a statistical estimate, and there is a small chance your true proportions are slightly different from that of the MLE, the MLE is the best estimate. If you want to keep it simple, simply use this MLE when describing your genetic heritage. If you need greater precision, more expensive tests can be performed (by our typing more Ancestry Informative Markers). Your MLE may shift slightly, and the confidence region around it would just shrink. An alternative to understanding more in terms of

your families heritage is to have more persons in your family (parents, grandparents, siblings) take the standard ***ANCESTRY 2.0*** test.

SEQUENCES FILE

In the sequences.doc file, we have provided you some of your genotypes within their natural chromosomal context, so that you can cut and paste them into various types of gene search engines. By doing this, you can learn about:

a) The human genome project
b) How our human DNA sequences are similar to, and differ from those of other organisms.
c) How cancer research is conducted through genetic research
d) And other things…

The sequences are just a few of those that we measure in order to determine your ancestral proportions. For any genetic position we measure, the variant sequence is flanked by invariant sequence (remember, 99.9% of our DNA is identical from person to person!). For example, if we measured whether you have a C or a T at position 25 in the following sequence:

GACCTACCATGATAAATTCCTAGG[C/T]GAGGAAGCTACTACACTGAGTTTAT

The site we actually sequence is indicated by the [C/T], and all of the other letters represent DNA nucleotides that are invariant. That is to say, they are the same from person to person. Since these invariant nucleotides have no impact on the estimate of your ancestral proportions, there is no reason to measure them. We provide you with this view of your sequences because you can use them to have fun and learn more about genomics at the same time – here is how.

The human genome project has spawned the development of myriad DNA databases, and you use your sequences in a search through these databases. We have all paid for the construction of these databases, and they are yours to use freely. Their main function is as a tool for scientists to better understand why it is that certain people are afflicted with diseases, or respond to drugs differently. All of us have a duty to educate ourselves so that we understand how this resource works, what information it provides and how it is used – because some of the information in these databases may very well save or extend your life some day and the wisest among us in the future will be able to critically evaluate the role of genomics and DNA in our health decisions.

Many of the databases can be reached through the website of our partner, Geospiza Corp. at http://www.geospiza.com/outreach/organelles/index.html. Geospiza operates this site as a

community service, and its main function is to help educate the layperson. In addition to searching with your sequences, you can click on links that will help you understand how the databases were constructed, and what information each offers.

For example, the links will take you to a site for you to search your sequences for relationship to those of other organisms, to identity whether any of the sequences have been flagged as over-"expressed" in various human cancers, or to learn from which chromosome and chromosomal region the sequences are targeted. If you are searching the cancer anatomy database, for example, you may refer to the genotypes.doc file and insert your particular genotype for any one sequence in the sequence.doc file (replacing the [C/T] with your sequences) and identify whether the variant sequence you harbor in your DNA is the same as that variant that is over-expressed in a certain type of cancer. Caution – this is not known to give you any information whatsoever about your predisposition to develop disease, but it is recreational information and as such, it is of some (as of yet unquantifiable) value. The possibilities for you to explore through this site are so numerous that we don't bother to list them all here – we invite you to be your own genomics private investigator and perform the searches you are interested in. We would be interested to learn what you find!

For a basic slide show introduction to genomics and DNA databases please visit:
http://www.geospiza.com/outreach/organelles/mutation/slide1.html

To screen your sequences against the human genome, all you need to do is copy and paste your sequences into the database query forms provided through the http://www.ncbi.nlm.nih.gov/ site (which can be accessed through the http://www.geospiza.com/outreach/organelles/index.html site). Click on the "BLAST" selection on the top menu and you have entered the human genome database search engine. Select "Standard nucleotide-nucleotide BLAST", and then simply paste your sequence (one sequence per query) into the Search box and click BLAST! A progress page will appear, and you will need to click the "Format!" Button to see your results. Learn more about the region of the chromosome corresponding to this sequence by clicking on the blue links…

Visit the Cancer Gene Anatomy Project at http://cgap.nci.nih.gov/, which is also accessible through the http://www.geospiza.com/outreach/organelles/index.html site. Here you can learn about the basics of biology and genes or do searches for certain genes that interest you. Find your favorite gene through http://cgap.nci.nih.gov/Genes/GeneFinder and learn more about it. Conduct your own genetic discovery experiments, and maybe you can be the one to figure out what makes cancer cells different from normal cells. Take one of your sequences and perform a BLAST search, as described above, to learn where the sequence is from (which chromosome, and region). Once you learn where this DNA comes from, go to http://gai.nci.nih.gov/html-

snp/ts.html and study the region around this segment of DNA. Are there any interesting genes nearby – for example, genes that are over-expressed in cancer? If so, the sequence region you are studying may also be over-expressed in cancer. Follow the various links to find out how you can search EST databases to find out whether this is the case. If it is the case, what is the variant form of the over-expressed gene and what is your variant? Cancer is much more complicated than one gene, and thought this type of research will provide you some information on your genetic background relative to data we currently have in our databases, it's value is strictly recreational and educational not diagnostic or therapeutic. We invite you to explore and learn, but don't get too carried away making inferences about your genetic background. First, the sites we have measured are not known to be linked with disease (though maybe one of you can prove us wrong on this statement through your research). It may be possible to someday identify a person at risk for cancer from their DNA, but that day is not yet upon us and when it is the tests for doing this will query those sequences the research has proven are linked to risk. Of the 2,000,000 SNPs in the genome, it is unlikely any of those are present in our panel…

The Scientific Foundation and validation of the test:

The science behind the test has been published in the scientific literature (for references, please see our web site at www.ancestrybydna.com). We have determined the frequency of DNA sequence variants in the various human populations (some of this data can be seen on the website), and by determining your sequence for each; we can determine the probability that you identify with each group. The test has been evaluated using a large number of people from a wide range of ancestral groups, and the estimates correspond well to what is known from anthropological and historical data. For example, Hispanics are known to have arisen as an ethnic group from the blending of colonial Europeans with Native Americans, and the hundreds of Hispanics we have tested align with these two groups almost exclusively, as expected. As another example, though most Nigerians plot as of unmixed African BioGeographical Ancestry (BGA), African Americans plot more as a mixture between this group and Europeans, which is also what would be expected from what we know about the admixture between Africans and Europeans in the US. The method has also been validated through pedigree challenge; when the BGA is determined from a mother and father, that of their children should plot somewhere between the two. To date, we have tested numerous family pedigrees, and the ancestral proportions of offspring always plot somewhere amongst those of their parents. When out MLE estimates are tested objectively (blindly), they prove to be excellent estimates of ancestral proportions.

When tested against known pedigrees, the test performs quite well. Some of the plots for the data below are shown on www.ancestrybydna.com website under "case study", where the data for DNAPrint's CSO is presented. Dr. Frudakis is a European American – his mother is European mix and his dad is mostly Greek. ANCESTRYbyDNA showed Dr. Frudakis to be of

mostly ancestry, but also of minority Native American ancestry. In fact, his paternal great-grandmother was either all or almost full-blood Cherokee. In addition to the unexpected Native American found in the mother's results, brings credence to the laws of genetics (we would predict a 12% Native American Ancestry for him if this great-grandmother was a 100% Native American and none of his other relatives were of Native American ancestry). His wife is Mexican, and she was determined to be of mostly Native American, but with some Native American and African heritage. This was also expected based on what we know from anthropological origin of the Hispanics (which were derived from the union of Spanish explorers, Native Americans, and West Africans in Colonial Caribbean and Latin America). Each of their 3 children plot roughly half way amongst both parents, as expected. None of the children shows any Asian/Pacific Islander ancestry, which would have been impossible (assuming an accurate test) since none of the parents showed any Asian/Pacific Islander heritage. Thus, the results of the children were consistent with those of the parents, and the MLE's are accurate estimates when tested against what is known from biographical data.

Experiment: Ancestrybydna 2.0 Blind trials on samples from families.
Purpose: To determine how well the test results agree with expectations formed from appreciation of a family pedigree.
Summary of Results:

Family 1						
Anc-1-Tony	INDO-EUROPEAN	92	EAST-ASIAN	0	NATIVE-AMERICAN	8
Anc-1-Wife	INDO-EUROPEAN	26	EAST-ASIAN	0	NATIVE-AMERICAN	74
Anc-1-TFC1	INDO-EUROPEAN	55	EAST-ASIAN	4	NATIVE-AMERICAN	41
Anc-1-TFC2	INDO-EUROPEAN	59	EAST-ASIAN	0	NATIVE-AMERICAN	41
Anc-1-TFC3	INDO-EUROPEAN	61	EAST-ASIAN	0	NATIVE-AMERICAN	39
Anc-1-PD_S-Data.INP	INDO-EUROPEAN	79	EAST-ASIAN	0	NATIVE-AMERICAN	21
Anc-1-PD_M-Data.INP	INDO-EUROPEAN	90	EAST-ASIAN	0	NATIVE-AMERICAN	10
Anc-1-PD_F-Data.INP	INDO-EUROPEAN	86	EAST-ASIAN	0	NATIVE-AMERICAN	14
Family 2						
Anc-1-PD_EG-Data.INP	INDO-EUROPEAN	79	EAST-ASIAN	21	NATIVE-AMERICAN	0
Anc-1-PD_MG-Data.INP	INDO-EUROPEAN	100	EAST-ASIAN	0	NATIVE-AMERICAN	0
Anc-1-PD_TG-Data.INP	INDO-EUROPEAN	80	EAST-ASIAN	20	NATIVE-AMERICAN	0
Anc-1-PD_VS-Data.INP	INDO-EUROPEAN	89	EAST-ASIAN	11	NATIVE-AMERICAN	0
Anc-1-PD_ZG-Data.INP	INDO-EUROPEAN	82	EAST-ASIAN	18	NATIVE-AMERICAN	0

Summary of Results:

Family 1: The father is F and mother is M. Both exhibit some Native American admixture (14% and 10%, respectively). Their children, Tony and S also both exhibit some Native American admixture (8% and 21%, respectively). Due to the law of independent assortment (the test markers span all 23 chromosomes), these results are reasonable. For example, if one of the children measured with 75% Native American, or 20% East Asian, these results would be unreasonable. Tony married his wife, who is Mexican (hence she types as 74% Native American), and they had three children, C1, C2, C3 Each of these three children TFC1, TFC3, TFC2 have 41%, 39% and 41% Native American ancestry, respectively, with the balance Indo-European. Again, these results are quite reasonable given the fact that the father was mostly Indo-European with slight Native American admixture and the mother was Hispanic, and mostly Native American with Indo-European admixture.

Family 2: For family two, the father is ZG and the mother is TG. Both have significant East Asian admixture. Their children are MG and EG. VS is the paternal grandmother. EG has 21% East Asian admixture, which is a reasonable amount considering the law of independent assortment. MG shows no detectable East Asian admixture, which is an unexpected result. For this child, it appears that we must invoke the 2%-5% error in order to explain the results. For example, it seems highly unlikely this child would have received no East Asian sequences from their parents (though possible). Rather, it is more likely she received a relatively small number from her parents, and the test was inadequate in detecting them.

For more information on how MLE's read for individuals of various mixed populations, please see www.ancestrybydna.com (product info section). You will find that the results are in good agreement with what is known from the anthropological history of each population and that all of the evidence we have to date suggests that ***ANCESTRYbyDNA 2.0*** test results are highly accurate.

Test Accuracy:

The genotypes (nucleotide letters) we have determined for you are quite accurate. Because we use the latest genetic reading equipment available, we routinely achieve a greater than 99% accuracy for each site. If an accurate value was not obtained for you at a particular site, you will see an "FL" instead of your letters for that site. Having a few of these generally does not prevent us from making a good ancestry estimate, but of course having too many would. Some reasons you may have an "FL" for a site include

a) A small region of your chromosome around this site is missing or is of different sequence character than for most. This result is not uncommon given the highly variable nature of the chromosomal positions we measure, and it certainly does not imply you have any sort of defect in any way whatsoever (in fact, it may be an indication of your uniqueness).

DNAPrint genomics, Inc.
900 Cocoanut Ave., Sarasota, FL 34236 tel 941-366-3400 fax 941-952-9770
www.ancestrybydna.com
email: admin@dnaprint.com.

b) We did not get enough DNA from your swab. The measurement of some markers is more sensitive to this than others. If there are too many "FLs" for your read-out, we will not be able to determine your ancestry proportions to a degree of accuracy that we would like, and in this case we will have asked you to submit another sample for a second try.

Experiment: Repeated estimation from the same samples (BD).

Purpose: To determine how reproducible the results are by measuring the proportions in the same individuals on different occasions. This measure will estimate genotyping error, handling error and algorithm error together as one error.

Generally speaking, in terms of reading your sequences correctly and in terms of error at the step of making the inference, again, your results are quite accurate. We have tested individuals on more than one occasion, and we see only a small, but acceptable level of error. For example, in the table below, we have performed repeats on 5 samples.

Results:

plate3-BD101-Data.INP	INDO-EUROPEAN	100	EAST-ASIAN	0	NATIVE-AMERICAN	0
plate5-BD101-Data.INP	INDO-EUROPEAN	100	EAST-ASIAN	0	NATIVE-AMERICAN	0
plate3-BD304-Data.INP	INDO-EUROPEAN	85	AFRICAN	15	NATIVE-AMERICAN	0
plate5-BD304-Data.INP	INDO-EUROPEAN	86	AFRICAN	14	NATIVE-AMERICAN	0
plate3-BD316-Data.INP	INDO-EUROPEAN	72	AFRICAN	27	EAST-ASIAN	1
plate5-BD316-Data.INP	INDO-EUROPEAN	79	AFRICAN	20	EAST-ASIAN	1
plate3-BD3162-Data.INP	INDO-EUROPEAN	100	EAST-ASIAN	0	NATIVE-AMERICAN	0
plate5-BD3162-Data.INP	INDO-EUROPEAN	89	EAST-ASIAN	5	NATIVE-AMERICAN	6
plate3-BD317-Data.INP	INDO-EUROPEAN	79	EAST-ASIAN	21	NATIVE-AMERICAN	0
plate5-BD317-Data.INP	INDO-EUROPEAN	84	EAST-ASIAN	16	NATIVE-AMERICAN	0

Summary of Results: The test shows a 5% error rate for the absolute percentage in any one group. Since this experiment we have been using 5 samples on each run as internal controls. The average error is 2-3% for these controls. Another group of 11 have also been tested repeatedly, and these show an average 2-3% error rate for 10 samples, and an average 5% error for the other sample. Best estimate from all of the data on repeated measurements is on order of

3-4% error rate for most determinations. Thus, if your profile came back as 96% Indo-European and 4% East Asian, it is debatable whether the 4% East Asian is significant.

How do the percentages align with physical appearance?

We have noted that individuals exhibiting physical characteristics of a population group generally have at least 25-30% identity with that group. For example, persons with an 85% Indo-European and 15% African generally exhibit few if any physical features characteristic of the African group, such as darker skin. Why is this? The genes that determine physical appearance are but a very small percentage of the total number of genes in the genome. Thus, for all of these genes to have sequences characteristic of one group, the person would need to be of relatively high proportions for that group. It's just the laws of probability. The higher the percentage of African a person is, the more likely the minority of the genome that determines physical appearance will be of African origin.

As a layperson, how can you better understand Human Genetics?
First of all, let us admit that genetics is a complex subject. You do not have to be a scientist to understand your results, and if you do not understand your results from thoroughly reading all of the materials supplied to you, you may want to do some basic reading on the internet.

The U.S. Department of Energy Human Genome Program is an excellent place to begin the process of understanding Human genetics.

http://www.ornl.gov/hgmis/publicat/primer2001/index.html

The following sites are also very informative for this purpose:

1) http://www.genome.gov/
2) www.sc.doe.gov/ober/hug_top.html
3) www.gene.ucl.ac.uk/hugo
4) www.wellcome.ac.uk
5) www.nhgri.nih.gov
6) www.doegenomestolife.org
7) www.ncbi.nlm.nih.gov/ncicgap
8) www.jgi.doe.gov
9) www.tigr.org

DNAPrint genomics, Inc.
900 Cocoanut Ave., Sarasota, FL 34236 tel 941-366-3400 fax 941-952-9770
www.ancestrybydna.com
email: admin@dnaprint.com.

For further reading, we recommend: The Great Human Diasporas; The History of Diversity and Evolution, by Cavalli-Sforza. It is an informative tool in understanding migration patterns and ancestral evolution.

People interested in discussing genealogy from DNA should visit this interactive forum:

http://archiver.rootsweb.com/th/index/GENEALOGY-DNA

Glossary.

Ancestry Informative Marker (AIM): AIMs are the subset of genetic markers that are different in allele frequencies across the populations of the world. Most polymorphism is shared among all populations and for most loci the most common allele is the same in each population.

Allele: Alternate sequences for a particular position in the genome. For example, a common variation in the genome is for some forms of the sequence to have Cytosine (C) while other forms have Thymidine (T). Thus, since we have two copies of each chromosome, there are three genotypes at this position CC, CT, and TT.

Chromosome: The physical units of heredity: long linear strands of DNA. Humans have 22 autosomal chromosome pairs, plus two sex chromosomes, X and Y. Men have two copies of each autosome, 1, 2, …, 22, X, Y. Women have two copies of each chromosome 1, 2, 3, …, 22, X, X. Each person thus has a total of 46 chromosomes.

Genomics: The study of the complete compliment of genetic material in a species.

Genome: All of the genetic material in a species. The human genome is approximately 3,300,000,000 base pairs in length.

Locus (pl. loci): The name for a physical position on the genome. Can either refer to a large region such as a complete gene or a very specific region, like a particular base pair position.

Polymorphism: The property of having more than one state or alternate sequence at a particular position. The alternate states are called alleles.

Single Nucleotide Polymorphism (SNP; pronounced snip): A precise base pair position where different people are found to vary in sequence. Generally two alternate alleles are found at a

particular SNP. At least 2,000,000 SNPs are now known and there may be over 30,000,000 in the human genome.

DNAPrint genomics, Inc.
900 Cocoanut Ave., Sarasota, FL 34236 tel 941-366-3400 fax 941-952-9770
www.ancestrybydna.com
email: admin@dnaprint.com.