The GenoChip: a new tool for genetic anthropology

S. Wells[1], E. Greenspan[2], S. Staats[2], T. Krahn[2], C. Tyler- Smith[3], Y. Xue[3], S. Tofanelli[4], P. Francalacci[5], F. Cucca[6], L. Pagani[7], L. Jin[8], H. Li[8], T.G. Schurr[9], J.B. Gaieski[9], C. Melendez[9], M.G. Vilar[9], A.C. Owings[9], R. Gomez[10], R. Fujita[11], F. Santos[12], D. Comas[13], O. Balanovsky[14], E. Balanovska[14], P. Zalloua[15], H. Soodyall[16], R. Pitchappan[17], G. Arun Kumar[17], M.F. Hammer[18], B. Greenspan[2], E. Elhaik[19]

1) Mission Programs, National Geographic Society, Washington, DC; 2) Family Tree DNA, Houston, TX; 3) Wellcome Trust Sanger Institute, Hinxton, UK; 4) University of Pisa, Italy; 5) University of Sassari, Italy; 6) National Research Council, Monserrato, Italy; 7) University of Cambridge, UK; 8) Fudan University, Shanghai, China; 9) University of Pennsylvania, Philadelphia, PA; 10) CINVESTAV, Mexico City, Mexico; 11) University of San Martin de Porres, Lima, Peru; 12) Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil; 13) Pompeu Fabra University, Barcelona, Spain; 14) Russian Academy of Medical Sciences, Moscow, Russia; 15) Lebanese American University, Byblos, Lebanon; 16) University of the Witwatersrand, Johannesburg, South Africa; 17) Chettinad Academy of Research and Education, Chennai, India; 18) University of Arizona, Tucson, AZ; 19) Johns Hopkins University, Baltimore, MD

INTRODUCTION

The Genographic Project is an international effort aimed at charting human history using genetic data. The project is non-profit and non-medical, and through the sale of its public participation kits it supports cultural preservation efforts in indigenous and traditional communities. To extend our knowledge of the human journey, interbreeding with ancient hominins, and modern human demographic history, we designed a genotyping chip optimized for genetic anthropology research.

OBJECTIVES

Our goal was to design, produce, and validate a SNP array dedicated to genetic anthropology. The GenoChip is an Illumina HD iSelect genotyping bead array with over 130,000 highly informative autosomal and X-chromosomal SNPs ascertained from over 450 worldwide populations (Figure 1), ~13,000 Y-chromosomal SNPs, and ~3,000 mtDNA SNPs. To determine the extent of gene flow from archaic hominins to modern humans, we included over 25,000 SNPs from candidate regions of interbreeding between extinct hominins (Neanderthal and Denisovan) and modern humans. To avoid any inadvertent medical testing we filtered out all SNPs that have known or suspected health or functional associations. We validated the chip by genotyping over 1,000 samples from 1000 Genomes, Family Tree DNA, and Genographic Project populations.

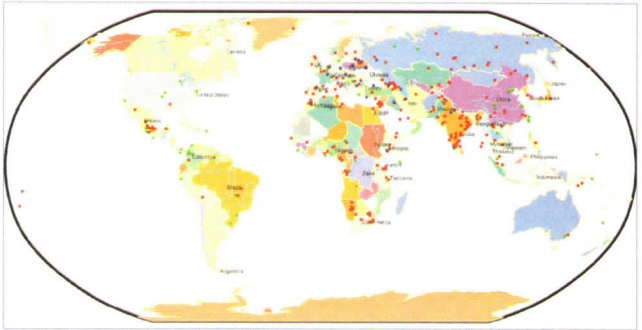

Figure 1. AIMs for 450 world populations were harvested from the literature (green) and from public and private collections (red) including 31 Jewish populations (blue).

The concordance between the GenoChip and the 1000 Genomes data was over 99.5%. The GenoChip has a SNP density of approximately (1/100,000) bases over 92% of the human genome (Figure 2) and is highly compatible with Illumina and Affymetrix commercial platforms (Table 1). The ~10,000 novel Y SNPs included on the chip have greatly refined our understanding of the Y-chromosome phylogenetic tree. By including Y and mtDNA SNPs on an unprecedented scale, the GenoChip is able to delineate extremely detailed human migratory paths (Figure 3). The autosomal and X-chromosomal markers included on the GenoChip have revealed novel patterns of ancestry that shed a detailed new light on human history (Figure 4). Interbreeding analysis with extinct hominins confirmed some previous reports and allowed us to describe the modern geographical distribution of these markers in detail.

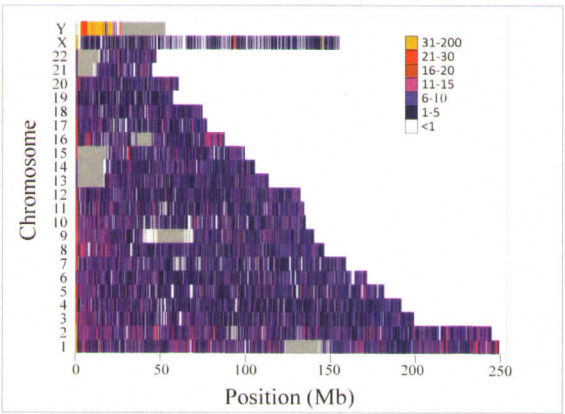

Figure 2. SNP density in the GenoChip. SNP densities across the genome are color coded to indicate the number of polymorphic SNPs per 100kb. Gaps in the assembly are shown as gray.

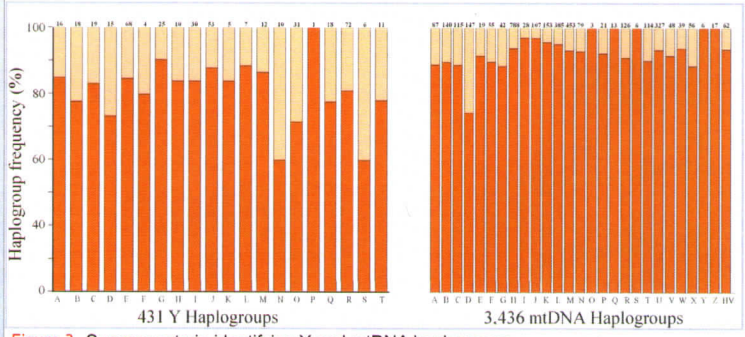

Figure 3. Success rate in identifying Y and mtDNA haplogroups.

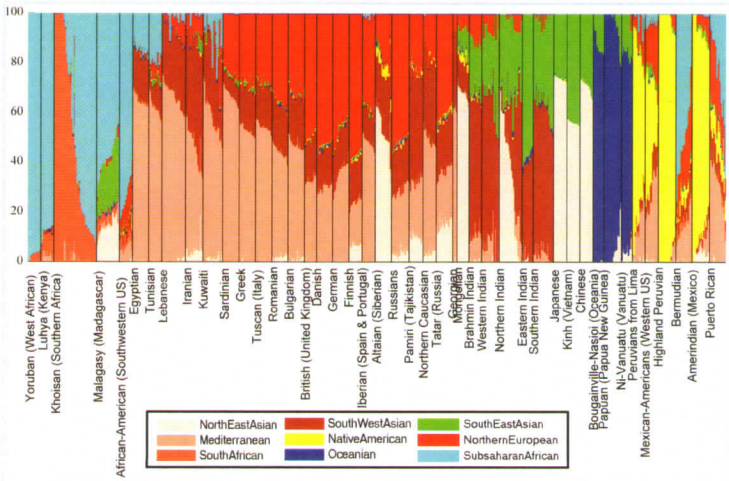

Figure 4. An unsupervised ADMIXTURE analysis (*K*=9) of 44 worldwide populations. In the legend, inferred regional affiliation of the ancestral components.

CONCLUSIONS

The GenoChip is the first genotyping chip completely dedicated to genetic anthropology with no known medically relevant markers. We anticipate that the large-scale application of the GenoChip using the Genographic Project's diverse sample collection will provide new insights into genetic anthropology and human history.

Platform	# Autosomes	Health related	Neanderthal	Denisovan	Chimpanzee	AIMs (Literature)	AIMs (AimsFinder)	Eskimo Saqqaq	Aboriginal	X chr haplotypes
GenoChip	130,000	-	23,962	1,357	998	49,930	66,966	10,159	12,027	975
Illumina quad	955,725	98,657	14,495	1,174	837	38,518	47,868	85,723	56,194	298
Illumina 2.5M	719,968	74,095	13,668	1,134	811	37,857	46,342	69,608	54,599	236
Illumina express	733,054	75,261	13,850	1,151	827	38,384	46,989	70,442	55,397	245
Illumina 660	544,366	56,845	13,524	1,533	1,105	60,574	70,253	49,670	81,136	208
Affy 5.0	443,327	79,868	8,394	261	190	47,459	35,317	29,470	13,034	142
Affy 6.0	908,195	112,435	16,012	541	403	52,146	48,259	61,429	29,116	459
Affy Human origins	482,934	236,537	9,962	781	576	45,319	33,324	50,233	25,561	31
23andMe (v3)	985,551	46,180	19,184	1,719	1,240	62,027	74,648	88,749	81,751	342
HapMap (phase3)	3,000,000	113,425	31,492	1,815	1,316	64,041	85,118	134,063	193,205	1,088

Table 1. Comparison of the number of SNPs in commercial and non-commercial chips over some of the categories used to design the GenoChip. *AimsFinder* is a PCA-based application designed to identify AIMS.

CONTACT

For more information on the Genographic Project: www.genographic.com

Corresponding author: Spencer Wells, spwells@ngs.org