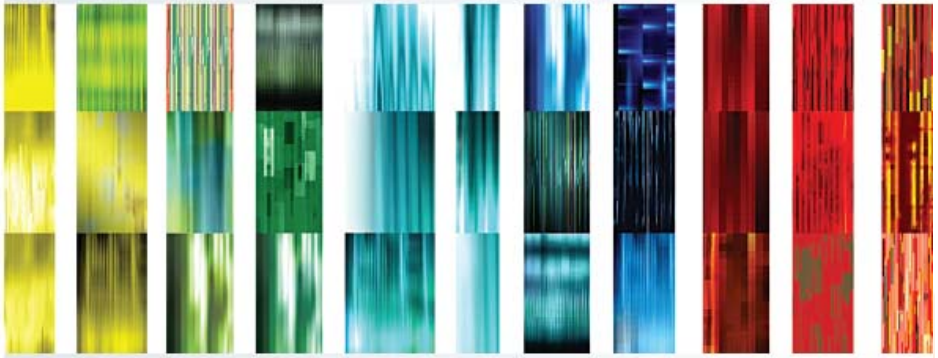
NAVIGENICS



6,750,000,000 – People in the world.

54,000,000 – Americans living with pre-diabetes.

20,000,000 – Americans who will have Alzheimer's disease in the year 2050.

7,500,000 – Americans suffering from heart attack.

4,000,000 – Americans who could avoid cancer by 2023 through lifestyle changes.

25,000–30,000 – Genes each person has in each cell of their body.

646 – Drug compounds in clinical trials for cancer treatment.

78 – Life expectancy in years for a baby born today.

33 – Percent of cardiovascular deaths that are premature.

16 – Percent of physician visits in U.S. related to prevention.

1 – person who can choose to use the power
of genetic information to improve their health.

Navigenics, Inc., - Founded 2006

- **Vision:**
- To **improve individuals' health** across the population by educating, empowering and motivating people to take action to prevent the onset of disease or lessen its impact.

- **Navigenics Health Compass:**
- Screens individuals for the *totality of their genetic risk* in a variety of common and rare diseases, and provides guidance and information on how avoid "environmental risk factors" in a focused way, engage in early screening efforts, and present earlier with disease – all of which are known to improve outcomes

I would like to communicate to you that …

- We are facing a health care crisis from Common Chronic Non-Infectious Disease (CCND) in this generation – and prevention is the only feasible solution
- Validated “genetic risk factors” are not so different than validated environmental risk factors
- Genetic risk factors can be used to refine risk in combination with other risk data and drive additional focused prevention behaviors and early detection paradigms
- Delivery of the information in an accurate and private fashion to the public is necessary to meet timelines

History and Expertise of Founders

Navigenics

Scientific Founders have Experience in WGA Studies

Common *Kibra* Alleles Are Associated with Human Memory Performance

Andreas Papassotiropoulos,[1,3*†] Dietrich A. Stephan,[3*†] Matthew J. Huentelman,[3] Frederic J. Hoerndli,[1] David W. Craig,[3] John V. Pearson,[3] Kim-Dung Huynh,[1] Fabienne Brunner,[1] Jason Corneveaux,[3] David Osborne,[4] M. Axel Wollmer,[1] Amanda Aerni,[1] Daniel Coluccia,[1] Jürgen Hänggi,[1] Christian R. A. Mondadori,[1] Andreas Buchmann,[1] Eric M. Reiman,[3,6] Richard J. Caselli,[5] Katharina Henke,[1] Dominique J.-F. de Quervain[1,2]

Human memory is a polygenic trait. We performed a genome-wide screen to identify memory-related gene variants. A genomic locus encoding the brain protein KIBRA was significantly associated with memory performance in three independent, cognitively normal cohorts from Switzerland and the United States. Gene expression studies showed that *KIBRA* was expressed in memory-related brain structures. Functional magnetic resonance imaging detected *KIBRA* allele–dependent differences in hippocampal activations during memory retrieval. Evidence from these experiments suggests a role for KIBRA in human memory.

Human memory is a polygenic cognitive trait. Heritability estimates of ~50% suggest that naturally occurring genetic variability has an important impact on this fundamental brain function (*1*). Recent candidate-gene association studies have identified some genetic variations with significant impact on human memory capacity (*2*–*5*). However, the success of these studies depends upon preexisting information, which limits their potential to identify unrecognized genes and molecular

ogeneity within the study sample (population structure) can lead to spurious associations between a genetic marker and a phenotype (*11*). Therefore, we controlled for genetic background and found no evidence of significant population stratification; the participants' genetic backgrounds formed one normally distributed cluster ($P = 0.6$) (*10*, *12*). We identified 10 participants as outliers (probability of cluster allocation lower than 25%) and excluded them from the genetic associ-

high statistical confidence (*10*). Two SNPs fulfilled these selection criteria and were prioritized for subsequent individual genotyping to exclude pooling-related false positives: rs17070145 and rs6439886. Both SNPs map within genes expressed in the human brain: rs17070145 is a common T → C substitution within the ninth intron of *KIBRA* (GenBank accession number NM_015238), encoding a neuronal protein, and rs6439886 is a common T → C substitution within the first intron of *CLSTN2* (encoding the synaptic protein calsyntenin 2) (NM_022131).

Both the *KIBRA* and *CLSTN2* SNPs were also significantly associated with differential human memory performance when we genotyped them individually in Swiss cohort 1 using an independent genotyping technology (*10*). Carriers of *KIBRA* rs17070145 T allele had 24% better free recall performance 5 min after word presentation ($P = 0.000004$) and 19% better free recall performance 24 hours after word presentation ($P = 0.0008$) than did noncarriers (Table 1, table S1, and fig. S2). TT and CT genotype groups of rs17070145 were combined because the frequency of the TT genotype was low and because both groups displayed similar memory performance (table S1). SNP rs6439886 yielded similar results; however, the mean difference of memory performance between geno-

Scientific Founders Experience in *Monogenic Disease*

European Journal of Human Genetics (2006) 14, 1097–1105

www.nature.com/ejhg

ARTICLE

A novel missense mutation in *ACTG1* causes dominant deafness in a Norwegian *DFNA20/26* family, but *ACTG1* mutations are not frequent among families with hereditary hearing impairment

Nanna D Rendtorff[1], Mei Zhu[2], Toril Fagerheim[3], Torben L Antal[4], MaryPat Jones[5], Tanya M Teslovich[6], Elizabeth M Gillanders[5], Michael Barmada[7], Erik Teig[8], Jeffrey M Trent[5,9], Karen H Friderici[2], Dietrich A Stephan[5,6,9] and Lisbeth Tranebjærg*[,1,3,10]

[1]*Department of Medical Biochemistry and Genetics, Wilhelm Johannsen Centre for Functional Genomics, University of Copenhagen, Copenhagen, Denmark;* [2]*Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, MI, USA;* [3]*Department of Medical Genetics, University Hospital, Tromsø, Norway;* [4]*The Kinase Signalling Laboratory, Biotech Research and Innovation Centre, Copenhagen, Denmark;* [5]*Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA;* [6]*Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA;* [7]*Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA, USA;* [8]*Department of Otorhinolaryngology, University Hospital of Oslo, Oslo, Norway;* [9]*Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ, USA;* [10]*Department of Audiology, H:S Bispebjerg Hospital, Copenhagen, Denmark*

The γ-actin gene (*ACTG1*) encodes a major cytoskeletal protein of the sensory hair cells of the cochlea. Recently, mutations in *ACTG1* were found to cause autosomal dominant, progressive, sensorineural hearing impairment linked to the *DFNA20/26* locus on chromosome 17q25.3 in four American families and in one Dutch family. We report here the linkage of autosomal dominant, progressive, sensorineural hearing impairment in a large Norwegian family to the *DFNA20/26* locus. Sequencing of *ACTG1* identified a novel missense mutation (c.1109T>C; p.V370A) segregating with the hearing loss. Functional analysis in yeast showed that the p.V370A mutation restricts cell growth at elevated temperature or under hyperosmolar stress. Molecular modelling suggested that the p.V370A mutation modestly alters a site for protein–protein interaction in γ-actin and thereby modestly alters γ-actin-based cytoskeletal structures. Nineteen Norwegian and Danish families with autosomal, dominant hearing impairment were analyzed for mutations in *ACTG1* by sequencing, but no disease-associated mutations were identified. Finally, a long-term follow-up of the hearing loss progression associated with the p.V370A mutation in *ACTG1* is provided. The present study expands our understanding of the genotype–phenotype relationship of this deafness gene and provides a sensitive and simple functional assay for missense mutations in this gene, which may assist future molecular diagnosis of autosomal-dominant hearing impairment. Finally, the present results do not indicate that mutations in *ACTG1* are a frequent cause of autosomal-dominant postlingual sensorineural hearing impairment in Norway nor Denmark.
European Journal of Human Genetics (2006) 14, 1097–1105. doi:10.1038/sj.ejhg.5201670; published online 14 June 2006

Keywords: *ACTG1*; *DFNA20/26*; deafness; hereditary; late-onset; progressive

*Correspondence: Dr L Tranebjærg, Department of Audiology, H:S Bispebjerg Hospital, Bispebjerg Bakke 23, DK-2400 Copenhagen NV, Denmark. Tel: +45 35 316341; Fax: +45 35 313951; E-mail: tranebjaerg@imbg.ku.dk
Received 9 November 2005; revised 24 April 2006; accepted 25 April 2006; published online 14 June 2006

Introduction

Non-syndromic hearing impairment (NSHI) is the most frequent sensory defect in humans and shows a very high degree of genetic heterogeneity.[1] More than 100 genes are

Human Molecular Genetics, 2001, Vol. 10, No. 3 189–194

Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2)

Natascia Tiso[1,+], Dietrich A. Stephan[2,+], Andrea Nava[3], Alessia Bagattin[1], Joseph M. Devaney[2,5], Fabio Stanchi[1], Gaelle Larderet[1], Bhoomi Brahmbhatt[2], Kevin Brown[2], Barbara Bauce[3], Michela Muriago[3], Cristina Basso[4], Gaetano Thiene[4], Gian Antonio Danieli[1,§] and Alessandra Rampazzo[1]

[1]Department of Biology, [3]Department of Cardiology and [4]Department of Pathology, University of Padova, 35121 Padova, Italy, [2]Research Centre for Genetic Medicine, Children's National Medical Centre, Washington, DC 20010, USA, [5]Transgenomics Inc., Gaithersburg, MD 20878, USA

Received 20 September 2000; Revised and Accepted 23 November 2000

Arrhythmogenic right ventricular dysplasia type 2 (ARVD2, OMIM 600996) is an autosomal dominant cardiomyopathy, characterized by partial degeneration of the myocardium of the right ventricle, electrical instability and sudden death. The disease locus was mapped to chromosome 1q42–q43. We report here on the physical mapping of the critical ARVD2 region, exclusion of two candidate genes (actinin 2 and nidogen), elucidation of the genomic structure of the cardiac ryanodine receptor gene (*RYR2*) and identification of *RYR2* mutations in four independent families. In myocardial cells, the RyR2 protein, activated by Ca^{2+}, induces the release of calcium from the sarcoplasmic reticulum into the cytosol. RyR2 is the cardiac counterpart of RyR1, the skeletal muscle ryanodine receptor, involved in malignant hyperthermia (MH) susceptibility and in central core disease (CCD). The RyR2 mutations detected in the present study occurred in two highly conserved regions, strictly corresponding to those where mutations causing MH or CCD are clustered in the *RYR1* gene. The detection of RyR2 mutations causing ARVD2, reported in this paper, opens the way to pre-symptomatic detection of carriers of the disease in childhood, thus enabling early monitoring and treatment.

INTRODUCTION

The acronym ARVD (arrhythmogenic right ventricular dysplasia) refers to a genetically heterogeneous group of cardiomyopathies characterized by progressive degeneration of the myocardium of the right ventricle, electrical instability and sudden death (1). This class of diseases, mostly inherited as autosomal dominant, is frequently involved in the cardiac sudden deaths of juveniles and athletes (2). Several forms with dominant inheritance (ARVD1, OMIM 1907970; ARVD2, 600996; ARVD3, 602086; ARVD4, 602087; ARVD5, 604400; and ARVD6, 604401) were identified in the last few years (3–8). Arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2) is clinically different from the other forms of ARVD because of the presence of peculiar effort-induced ventricular arrhythmias, by its high penetrance and by a 1:1 male:female ratio among the affected subjects. This clinical entity was first described in 1988 by Nava *et al.* (9), when juvenile sudden death and effort-induced ventricular tachycardias were reported in a family with right ventricular cardiomyopathy. The ARVD2 disease locus was mapped to chromosome 1q42–q43 (4,10). We report here on the physical mapping of the critical ARVD2 region, including radiation hybrid placement of candidate genes, elucidation of their genomic structure and identification of cardiac ryanodine receptor (RYR2) missense mutations in four independent ARVD2 families. The pathogenetic role of the detected mutations is discussed.

RESULTS AND DISCUSSION

Four families showing recurrence of ARVD2 cases (Fig. 1) were detected in the course of an investigation on ARVD families, lasting over a decade. The clinical description of families 102 and 115, both from Venetia (north-east Italy), was reported elsewhere (4,10); families 122 and 123, recently recruited to the study, came from Lombardy and Venetia, respectively. In all cases, the clinical phenotype showed neither inter- nor intrafamilial variability.

The pathological trait was inherited as autosomal dominant and linkage analysis proved that the disease was inherited linked to markers *ACTN2* (CA4F/R) and *D1S2680* (data not shown).

By radiation hybrid mapping and sequence tagged site (STS) content, we mapped three ARVD2 candidate genes to the ARVD2 critical interval: α actinin-2 (*ACTN2*), nidogen (*NID*)

+These authors contributed equally to this work
§To whom correspondence should be addressed. Tel: +39 049 8276215; Fax: +39 049 8276209; Email: danieli@bio.unipd.it

Scientific Founders Experience in *Monogenic Disease*

The NEW ENGLAND JOURNAL of MEDICINE

BRIEF REPORT

Recessive Symptomatic Focal Epilepsy and Mutant Contactin-Associated Protein-like 2

Kevin A. Strauss, M.D., Erik G. Puffenberger, Ph.D., Matthew J. Huentelman, Ph.D., Steven Gottlieb, M.D., Seth E. Dobrin, Ph.D., Jennifer M. Parod, B.S., Dietrich A. Stephan, Ph.D., and D. Holmes Morton, M.D.

SUMMARY

Contactin-associated protein-like 2 (CASPR2) is encoded by *CNTNAP2* and clusters voltage-gated potassium channels ($K_v1.1$) at the nodes of Ranvier. We report a homozygous mutation of *CNTNAP2* in Old Order Amish children with cortical dysplasia, focal epilepsy, relative macrocephaly, and diminished deep-tendon reflexes. Intractable focal seizures began in early childhood, after which language regression, hyperactivity, impulsive and aggressive behavior, and mental retardation developed in all children. Resective surgery did not prevent the recurrence of seizures. Temporal-lobe specimens showed evidence of abnormalities of neuronal migration and structure, widespread astrogliosis, and reduced expression of CASPR2.

From the Clinic for Special Children, Strasburg, Pa. (K.A.S., E.G.P., D.H.M.); the Translational Genomics Research Institute, Phoenix, Ariz. (M.J.H., J.M.P., D.A.S.); Lancaster General Hospital, Lancaster, Pa. (S.G.); and the Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, Wis. (S.E.D.). Address reprint requests to Dr. Strauss at the Clinic for Special Children, 535 Bunker Hill Rd., Strasburg, PA 17579, or at kstrauss@clinicforspecialchildren.org or to Dr. Stephan at the Translational Genomics Research Institute, 445 N. 5th St., Phoenix, AZ 85004, or at dstephan@tgen.org.

*Drs. Strauss and Puffenberger contributed equally to this article.

N Engl J Med 2006;354:1370-7.

MOST EPILEPTIC DISORDERS CAN BE TRACED TO AN ABNORMALITY OF cortical architecture, channel-mediated currents, neuronal growth and differentiation, or cerebral metabolism.[1,2] In most cases, however, the underlying biologic complexity of epilepsy precludes the identification of the genetic cause, and 65 to 79 percent of recurrent seizure syndromes remain unexplained.[3] Microarray analysis of DNA samples can be a powerful tool for revealing a genetic lesion in well-defined families. We have used this approach in Old Order Amish families, some members of which have a clinical and neuropathological phenotype that we designate as the cortical dysplasia–focal epilepsy (CDFE) syndrome. We identified a genetic variation in the gene encoding CASPR2 in affected patients, a finding that suggests that CASPR2 influences brain development.

METHODS

The study was approved by the Western Institutional Review Board of Olympia, Washington, and written informed consent was obtained from all participating parents. Phenotype information is based on clinical data from nine patients between the ages of two and nine years. Clinical investigations were routine. Methods are described briefly here; details are included in the Supplementary Appendix, which is available with the full text of this article at www.nejm.org.

Four affected children and their six parents were used for analysis of single-nucleotide polymorphisms (SNPs) with the use of the GeneChip Human Mapping 10K assay kit (Affymetrix). Genotype data were analyzed with Varia software (Silicon Genetics), which assumes mutation homogeneity and scans for regions that are autozygous (identical by descent) among affected persons. Target gene sequencing was performed as previously described.[4]

Serial 8-μm sections from paraffin-embedded temporal-lobe specimens were

Mapping of sudden infant death with dysgenesis of the testes syndrome (SIDDT) by a SNP genome scan and identification of *TSPYL* loss of function

Erik G. Puffenberger*†, Diane Hu-Lince†‡, Jennifer M. Parod†‡, David W. Craig‡, Seth E. Dobrin‡, Andrew R. Conway§, Elizabeth A. Donarum¶, Kevin A. Strauss*, Travis Dunckley‡, Javier F. Cardenas‡, Kara R. Melmed‡, Courtney A. Wright‡, Winnie Liang‡, Phillip Stafford‡, C. Robert Flynn‖, D. Holmes Morton†, and Dietrich A. Stephan‡**

*Clinic for Special Children, Strasburg, PA 17579; ‡Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ 85004; §Silicon Genetics, Redwood City, CA 94063; ¶Department of Neurodevelopmental Genetics, Barrow Neurological Institute, Phoenix, AZ 85012; and ‖Arizona BioDesign Institute and Harrington Department of Bioengineering, Arizona State University, Tempe, AZ 85248

Edited by Albert de la Chapelle, Ohio State University, Columbus, OH, and approved June 15, 2004 (received for review February 19, 2004)

We have identified a lethal phenotype characterized by sudden infant death (from cardiac and respiratory arrest) with dysgenesis of the testes in males [Online Mendelian Inheritance in Man (OMIM) accession no. 608800]. Twenty-one affected individuals with this autosomal recessive syndrome were ascertained in nine separate sibships among the Old Order Amish. High-density single-nucleotide polymorphism (SNP) genotyping arrays containing 11,555 single-nucleotide polymorphisms evenly distributed across the human genome were used to map the disease locus. A genome-wide autozygosity scan localized the disease gene to a 3.6-Mb interval on chromosome 6q22.1-q22.31. This interval contained 27 genes, including two testis-specific Y-like genes (*TSPYL* and *TSPYL4*) of unknown function. Sequence analysis of the *TSPYL* gene in affected individuals identified a homozygous frameshift mutation (457.458insG) at codon 153, resulting in truncation of translation at codon 169. Truncation leads to loss of a peptide domain with strong homology to the nucleosome assembly protein family. GFP-fusion expression constructs were constructed and illustrated loss of nuclear localization of truncated TSPYL, suggesting loss of a nuclear localization patch in addition to loss of the nucleosome assembly domain. These results shed light on the pathogenesis of a disorder of sexual differentiation and brainstem-mediated sudden death, as well as give insight into a mechanism of transcriptional regulation.

GENETICS

Over two generations, nine families from the Belleville Amish Community have lost 21 infants to a recently discovered disorder we have entitled sudden infant death with dysgenesis of the testes syndrome (SIDDT; Fig. 1) [Online Mendelian Inheritance in Man (OMIM) accession no. 608800]. The condition is not seen in the Lancaster County Old Order Amish population. Three of these infants have been cared for at the Clinic for Special Children. Most previous cases were studied extensively at regional medical centers; however, no diagnostic tests were available, and clinical recognition of the syndrome was difficult, particularly in affected females. Caretakers say that they can often recognize affected infants at birth by the unusual sound of their cry, which is a staccato sound, similar to the cry of a goat (see *Materials and Methods* for a complete description of the syndrome).

Homozygosity mapping to identify disease genes in autosomal recessive disorders common in founder populations by using traditional methods has often been hampered by microsatellite marker density (1, 2). Typically, ≈400 microsatellite markers are used in such linkage studies spaced at an average 10-centimorgan density throughout the genome. Single-nucleotide polymorphisms (SNP) are present in the genome at an average density of ≈1 per 1,300 bp and hold enormous potential as a high-density high-throughput genotyping strategy for disease gene mapping (3). Information content is a function of marker heterozygosity, distance between markers, and pedigree structure. Although each individual biallelic SNP is less informative than a single microsatellite marker in most cases (e.g., average heterozygosity of 0.37 on Affymetrix 10K Array vs. 0.72 on the Center for Inherited Disease Research database), the greater number of SNPs in aggregate leads to higher information content at any particular point in the genome (4). The Affymetrix 10K Array assay requires only 250 ng of DNA and generates 11,555 SNP genotypes with an average resolution of one SNP every 210 kb (5). This new genotyping platform has a throughput 100-fold greater than microsatellite genotyping and an accuracy of >99.5% with automated allele calling. The high density and information content of this genotyping platform make it ideal for localizing small regions of homozygosity.

Blinded validation studies were done to verify that gene mapping could be accomplished using the SNP arrays in small disease pedigrees with known map location. The gene for each disease was previously mapped and the causative mutation identified. In each case, we were able to reproduce the mapping results. A software analysis package entitled VARIA (Silicon Genetics) was developed to handle the large amount of data inherent to these assays and to correctly localize disease-carrying loci from markers that are in linkage disequilibrium with one another. The genome-wide linkage scan conducted on the multiplex SIDDT pedigree rapidly and unambiguously mapped this disorder to 6q22 with a location score of 8.11 [maximum 2-point logarithm of odds (LOD) of 2.41] in a 3.6-Mb interval. Sequencing of two candidate genes in the region identified a nonsense mutation in the testis-specific Y-like gene (*TSPYL*). Functional validation was performed through construction of GFP-fusion proteins of both the full-length and truncated TSPYL proteins to investigate the effect of truncation on cellular localization.

Materials and Methods

Subjects and Samples. DNA samples used in mapping and sequencing studies of SIDDT were acquired by the Clinic for Special Children, with informed consent. Peripheral blood was collected from four affected individuals, their parents, siblings, and extended family members. Samples from maple syrup urine

This paper was submitted directly (Track II) to the PNAS office.

Freely available online through the PNAS open access option.

Abbreviations: SIDDT, sudden infant death with dysgenesis of the testes syndrome; TSPYL, testis-specific Y-like; SNP, single-nucleotide polymorphism; NAP, nucleosome assembly protein; LOD, logarithm of odds.

Data deposition: The disorder reported in this paper has been deposited in the Online Mendelian Inheritance in Man (OMIM) database (accession no. 608800).

†E.G.P., D.H.-L, and J.M.P. contributed equally to this work.

**To whom correspondence should be addressed. E-mail: dstephan@tgen.org.

Navigenics

HGP Enables WGA Studies in Common Disease

- HapMap Project allows us to use representative SNPs to screen through the genome and identify areas which are enriched in persons with disease

Scientific Founders Experience in *Complex Disease*

Neuron
Report

GAB2 Alleles Modify Alzheimer's Risk in *APOE* ε4 Carriers

Eric M. Reiman,[1,2,3,17,18,*] Jennifer A. Webster,[1,17,18] Amanda J. Myers,[4,5,18] John Hardy,[5,6] Travis Dunckley,[1,17] Victoria L. Zismann,[1,17] Keta D. Joshipura,[1,17] John V. Pearson,[1,17] Diane Hu-Lince,[1,17] Matthew J. Huentelman,[1,17] David W. Craig,[1,17] Keith D. Coon,[1,7,17] Winnie S. Liang,[1,17] RiLee H. Herbert,[1,17] Thomas Beach,[8,17] Kristen C. Rohrer,[5] Alice S. Zhao,[5] Doris Leung,[5] Leslie Bryden,[5] Lauren Marlowe,[5] Mona Kaleem,[5] Diego Mastroeni,[8] Andrew Grover,[8,17] Christopher B. Heward,[9] Rivka Ravid,[10] Joseph Rogers,[8,17] Michael L. Hutton,[11] Stacey Melquist,[11] Ron C. Petersen,[12] Gene E. Alexander,[13,17] Richard J. Caselli,[14,17] Walter Kukull,[16] Andreas Papassotiropoulos,[1,15] and Dietrich A. Stephan[1,2,17,*]

[1] Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ, 85004, USA
[2] Banner Alzheimer's Institute, Phoenix, AZ 85006, USA
[3] Department of Psychiatry, University of Arizona, Tucson, AZ 85724, USA
[4] Department of Psychiatry and Behavioral Sciences, University of Miami, Miller School of Medicine, Miami, FL 33136, USA
[5] Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, 20892, USA
[6] Reta Lila Weston Laboratories, Department of Molecular Neuroscience, Institute of Neurology, Queen Square, London WC1N, 3BG, England
[7] Division of Thoracic Oncology Research, St. Joseph's Hospital, Phoenix, AZ 85013, USA
[8] Sun Health Research Institute, Sun City, AZ 85351, USA
[9] Kronos Science Laboratory, Phoenix, AZ 85016, USA
[10] Netherlands Institute for Neurosciences, Dutch Royal Academy of Arts and Sciences, Meibergdreef 47 AB Amsterdam, The Netherlands
[11] Department of Neuroscience, Mayo Clinic, Jacksonville, FL 32224, USA
[12] Department of Neurology, Mayo Clinic, Rochester, MN 55905, USA
[13] Department of Psychology, Arizona State University, Tempe, AZ 85281, USA
[14] Department of Neurology, Mayo Clinic, Scottsdale, AZ 85259, USA
[15] Division of Molecular Psychology and Life Sciences Training Facility, Biozentrum, University of Basel, Switzerland
[16] National Alzheimer's Coordinating Center, Department of Epidemiology, School of Public Health and Community Medicine,

Scientific Founders Experience in *Complex Disease*

A High-Density Whole-Genome Association Study Reveals That *APOE* Is the Major Susceptibility Gene for Sporadic Late-Onset Alzheimer's Disease

Keith D. Coon, Ph.D.; Amanda J. Myers, Ph.D.; David W. Craig, Ph.D.; Jennifer A. Webster, B.A.; John V. Pearson, B.Sc.; Diane Hu Lince, Ph.D.; Victoria L. Zismann, M.S.; Thomas G. Beach, M.D.; Doris Leung, M.D.; Leslie Bryden, M.S.; Rebecca F. Halperin, B.Sc.; Lauren Marlowe, B.Sc.; Mona Kaleem, B.Sc.; Douglas G. Walker, Ph.D.; Rivka Ravid, Ph.D.; Christopher B. Heward, Ph.D.; Joseph Rogers, Ph.D.; Andreas Papassotiropoulos, M.D.; Eric M. Reiman, M.D.; John Hardy, Ph.D.; and Dietrich A. Stephan, Ph.D.

Objective: While the apolipoprotein E (*APOE*) ε4 allele is a well-established risk factor for late-onset Alzheimer's disease (AD), initial genome scans using microsatellite markers in late-onset AD failed to identify this locus on chromosome 19. Recently developed methods for the simultaneous assessment of hundreds of thousands of single nucleotide polymorphisms (SNPs) promise to help more precisely identify loci that contribute to the risk of AD and other common multigenic conditions. We sought here to demonstrate that more precise identification of loci that are associated with complex, multigenic genetic disorders can be achieved using ultra-high-density whole-genome associations by demonstrating their ability to identify the *APOE* locus as a major sus-

Received Oct. 20, 2006; accepted Jan. 30, 2007. *From the Neurogenomics Division, Translational Genomics Research Institute, Phoenix, Ariz. (Drs. Coon, Craig, Lince, Papassotiropoulos, Reiman, and Stephan; Mss. Webster, Zismann, and Halperin; and Mr. Pearson); the Department of Psychiatry and Behavioral Sciences, University of Miami, Miller School of Medicine, Miami, Fla. (Dr. Myers); the Sun Health Research Institute, Sun City, Ariz. (Drs. Beach, Walker, and Rogers); the Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Md. (Drs. Leung and Hardy and Mss. Bryden, Marlowe, and Kaleem); the Royal Dutch Academy of Sciences, Amsterdam, the Netherlands (Dr. Ravid); Kronos Science Laboratories, Phoenix, Ariz. (Dr. Heward); the Division of Psychiatry Research, University of Zurich, Zurich, Switzerland (Dr. Papassotiropoulos); and Banner Alzheimer's Institute, Phoenix (Dr. Reiman); the Department of Psychiatry, University of Arizona, Tucson (Dr. Reiman); and Arizona Alzheimer's Consortium, Phoenix (Drs. Reiman and Stephan), Ariz.*

Support statements and acknowledgments appear at the end

Scientific Founders Experience in *Complex Disease*

Original Paper

Neuro–degenerative Diseases

Neurodegenerative Dis
DOI: 10.1159/000110789

Received: March 8, 2007
Accepted after revision: June 4, 2007
Published online: November 1, 2007

Sorl1 as an Alzheimer's Disease Predisposition Gene?

Jennifer A. Webster[a] Amanda J. Myers[b] John V. Pearson[a] David W. Craig[a] Diane Hu-Lince[a] Keith D. Coon[a] Victoria L. Zismann[a] Thomas Beach[c, d] Doris Leung[e] Leslie Bryden[e] Rebecca F. Halperin[a] Lauren Marlowe[e] Mona Kaleem[e] Matthew J. Huentelman[a] Keta Joshipura[a] Douglas Walker[c, d] Christopher B. Heward[f] Rivka Ravid[i] Joseph Rogers[c, d] Andreas Papassotiropoulos[a, j] John Hardy[b] Eric M. Reiman[a, d, g, h] Dietrich A. Stephan[a, d]

[a]Neurogenomics Division, Translational Genomics Research Institute, Phoenix, Ariz.; [b]Department of Psychiatry and Behavioral Sciences, University of Miami, Miller School of Medicine, Miami, Fla.; [c]Sun Health Research Institute, Sun City, Ariz.; [d]Arizona Alzheimer's Consortium, Phoenix, Ariz.; [e]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Md.; [f]Kronos Science Laboratories, [g]Banner Alzheimer's Institute, Phoenix, Ariz., and [h]Department of Psychiatry, University of Arizona, Tucson, Ariz., USA; [i]Royal Dutch Academy of Sciences, Amsterdam, The Netherlands, and [j]Division of Psychiatry Research, University of Zurich, Zurich, Switzerland

Scientific Founders Experience in *Complex Disease*

BRIEF COMMUNICATIONS

nature genetics

Common sequence variants on 20q11.22 confer melanoma susceptibility

Kevin M Brown[1,23], Stuart MacGregor[2,23], Grant W Montgomery[2], David W Craig[3], Zhen Zhen Zhao[2], Kelly Iyadurai[1], Anjali K Henders[2], Nils Homer[4], Megan J Campbell[2], Mitchell Stark[2], Shane Thomas[2], Helen Schmid[5], Elizabeth A Holland[5], Elizabeth M Gillanders[6], David L Duffy[2], Judith A Maskiell[7], Jodie Jetann[8], Megan Ferguson[8], Dietrich A Stephan[3], Anne E Cust[7], David Whiteman[2], Adele Green[2], Håkan Olsson[9,21,22], Susana Puig[10,22], Paola Ghiorzo[11,22], Johan Hansson[12,22], Florence Demenais[13,22], Alisa M Goldstein[14], Nelleke A Gruis[15,22], David E Elder[16,22], Julia Newton Bishop[17,22], Richard F Kefford[5], Graham G Giles[18], Bruce K Armstrong[19], Joanne F Aitken[8], John L Hopper[7], Nicholas G Martin[2], Jeffrey M Trent[20], Graham J Mann[5] & Nicholas K Hayward[2]

We conducted a genome-wide association pooling study for

have been identified (*CDKN2A*, *ARF*, *CDK4* and a locus on 1p22)[2], and *MC1R* has been validated as a gene harboring low-penetrance risk alleles[3,4].

To identify additional low-penetrance risk alleles, we carried out a genome-wide association study (GWAS) involving the pooling of 864 cases drawn from a larger population-based sample of cases (individuals with melanoma) from Queensland, unselected for age at onset (Queensland study of Melanoma: Environment and Genetic Associations (Q-MEGA)[5]), and 864 controls (Q1). Each pool was hybridized to six Illumina HumanHap550 arrays, and SNPs were ranked after accounting for pooling error[6,7]. The proportion of SNPs with P values from pooling of < 0.01 was consistent with what would be expected by chance if there were no true associations. Conversely, at smaller P-value thresholds, there were more SNPs than expected by chance. For example, at the 0.0001 threshold, we would expect to see ~55 SNPs under the null hypothesis of no association, but we in fact observed 90 SNPs, indicating that there were a number of true associations (**Supplementary Note** online).

Here we focus on only the most significant finding from pooling. The first-ranked (rs17305657, $P = 2.56 \times 10^{-7}$) and fourth-ranked (rs4911442, $P = 2.39 \times 10^{-6}$) SNPs are 1.5 Mb apart on chromosome

Scientific Founders Experience in *Complex Disease*

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Whole-Genome Analysis of Sporadic Amyotrophic Lateral Sclerosis

Travis Dunckley, Ph.D., Matthew J. Huentelman, Ph.D., David W. Craig, Ph.D., John V. Pearson, B.Sc., Szabolcs Szelinger, B.S., Keta Joshipura, B.S., Rebecca F. Halperin, B.Sc., Chelsea Stamper, B.S., Kendall R. Jensen, Ph.D., David Letizia, M.S., Sharon E. Hesterlee, Ph.D., Alan Pestronk, M.D., Todd Levine, M.D., Tulio Bertorini, M.D., Michael C. Graves, M.D., Tahseen Mozaffar, M.D., Carlayne E. Jackson, M.D., Peter Bosch, M.D., April McVey, M.D., Arthur Dick, M.D., Richard Barohn, M.D., Catherine Lomen-Hoerth, M.D., Jeffrey Rosenfeld, M.D., Daniel T. O'Connor, M.D., Kuixing Zhang, M.D., Ph.D., Richard Crook, Ph.D., Henrik Ryberg, Ph.D., Michael Hutton, Ph.D., Jonathan Katz, M.D., Ericka P. Simpson, M.D., Hiroshi Mitsumoto, M.D., Robert Bowser, Ph.D., Robert G. Miller, M.D., Stanley H. Appel, M.D., and Dietrich A. Stephan, Ph.D.

ABSTRACT

BACKGROUND

Approximately 90% of persons with amyotrophic lateral sclerosis (ALS) have the sporadic form, which may be caused by the interaction of multiple environmental factors and previously unknown genes.

METHODS

From the Translational Genomics Research Inst., Phoenix, AZ (T.D., M.J.H., D.W.C., J.V.P., S.S., K.J., R.F.H., C.S., K.R.J., D.L., D.A.S.); Muscular Dystrophy Association, Tucson, AZ (S.E.H.); Washington Univ. School of Medicine, St. Louis (A.P.);

Scientific Founders Experience in *Complex Disease*

AJHG, 2007

ARTICLE

Identification of the Genetic Basis for Complex Disorders by Use of Pooling-Based Genomewide Single-Nucleotide–Polymorphism Association Studies

John V. Pearson,* Matthew J. Huentelman,* Rebecca F. Halperin, Waibhav D. Tembe, Stacey Melquist, Nils Homer, Marcel Brun, Szabolcs Szelinger, Keith D. Coon, Victoria L. Zismann, Jennifer A. Webster, Thomas Beach, Sigrid B. Sando, Jan O. Aasly, Reinhard Heun, Frank Jessen, Heike Kölsch, Magdalini Tsolaki, Makrina Daniilidou, Eric M. Reiman, Andreas Papassotiropoulos, Michael L. Hutton, Dietrich A. Stephan, and David W. Craig

We report the development and validation of experimental methods, study designs, and analysis software for pooling-based genomewide association (GWA) studies that use high-throughput single-nucleotide–polymorphism (SNP) genotyping microarrays. We first describe a theoretical framework for establishing the effectiveness of pooling genomic DNA as a low-cost alternative to individually genotyping thousands of samples on high-density SNP microarrays. Next, we describe software called "GenePool," which directly analyzes SNP microarray probe intensity data and ranks SNPs by increased likelihood of being genetically associated with a trait or disorder. Finally, we apply these methods to experimental case-control data and demonstrate successful identification of published genetic susceptibility loci for a rare monogenic disease (sudden infant death with dysgenesis of the testes syndrome), a rare complex disease (progressive supranuclear palsy), and a common complex disease (Alzheimer disease) across multiple SNP genotyping platforms. On the basis of these theoretical calculations and their experimental validation, our results suggest that pooling-based GWA studies are a logical first step for determining whether major genetic associations exist in diseases with high heritability.

Genomewide association (GWA) studies that use hundreds of thousands of SNPs have the potential to revolutionize our ability to identify the genetic influences of complex traits and diseases. Although potentially allowing for the identification of common variants to complex disease, GWA studies often require millions of dollars to complete and, as such, are beyond the reach of many research

of pooling on SNP genotyping microarrays (or related technologies). With a few exceptions, these reports have focused on predicting allelic frequencies across thousands of SNPs rather than on the effectiveness of pooling in identifying the genetic basis of complex disorders.[1–21] Indeed, it is not yet clear whether predicting allelic frequency to within 2% accuracy (as is frequently reported)

Scientific Founders Experience in *Complex Disease*

LETTERS

nature genetics

A survey of genetic human cortical gene expression

Amanda J Myers[1,2,10], J Raphael Gibbs[1,3,10], Jennifer A Webster[4,5,10], Kristen Rohrer[1], Alice Zhao[1], Lauren Marlowe[1], Mona Kaleem[1], Doris Leung[1], Leslie Bryden[1], Priti Nath[1], Victoria L Zismann[4,5], Keta Joshipura[4,5], Matthew J Huentelman[4,5], Diane Hu-Lince[4,5], Keith D Coon[4–6], David W Craig[4,5], John V Pearson[4,5], Peter Holmans[7], Christopher B Heward[8], Eric M Reiman[4,5,9], Dietrich Stephan[4,5,9] & John Hardy[1,3]

It is widely assumed that genetic differences in gene expression underpin much of the difference among individuals and many of the quantitative traits of interest to geneticists. Despite this, there has been little work on genetic variability in human gene expression and almost none in the human brain, because tools for assessing this genetic variability have not been available. Now, with whole-genome SNP genotyping arrays and whole-transcriptome expression arrays, such experiments have become feasible. We have carried out whole-genome

not receive any neurologic assessment. Very little has been done with other tissues because of their inaccessibility. However, it is well established that mRNA is stable postmortem in the human brain[8], and our and others' studies have shown that the apolipoprotein E (*APOE*) and microtubule-associated protein tau (*MAPT*) genes are subject to distortions in allelic expression[9–11]. Additionally, several studies using inbred mouse strains have mapped important expression quantitative trait loci (eQTL) in the mouse brain[1–3]. With this background, we developed a resource that allows the assessment of the

Scientific Founders Experience in *Complex Disease*

BIOINFORMATICS

Vol. 00 no. 0 2005, pages 1–5
doi:10.1093/bioinformatics/bti283

Gene expression

SNiPer-HD: Improved Genotype Calling Accuracy by an Expectation-Maximization Algorithm for High-Density SNP Arrays

Jianping Hua[1], David W. Craig[2], Marcel Brun[1], Jennifer Webster[2], Victoria Zismann[2], Waibhav Tembe[1], Keta Joshipura[2], Matthew J. Huentelman[2], Edward R. Dougherty[1,3] and Dietrich A. Stephan[2*]

[1]Computation Biology Division and [2]Neurogenomics Division, Translational Genomics Research Institute, Phoenix, 445 N 5th st., Phoenix, AZ and [3]Department of Electrical & Computer Engineering, Texas A&M University, College Station, TX

ABSTRACT

Motivation: The technology to genotype single nucleotide polymorphisms (SNPs) at extremely high densities provides for hypothesis-free genome-wide scans for common polymorphisms associated with complex disease. However, we find that errors introduced by commonly employed genotyping algorithms may lead to inflation of false associations between markers and phenotype.

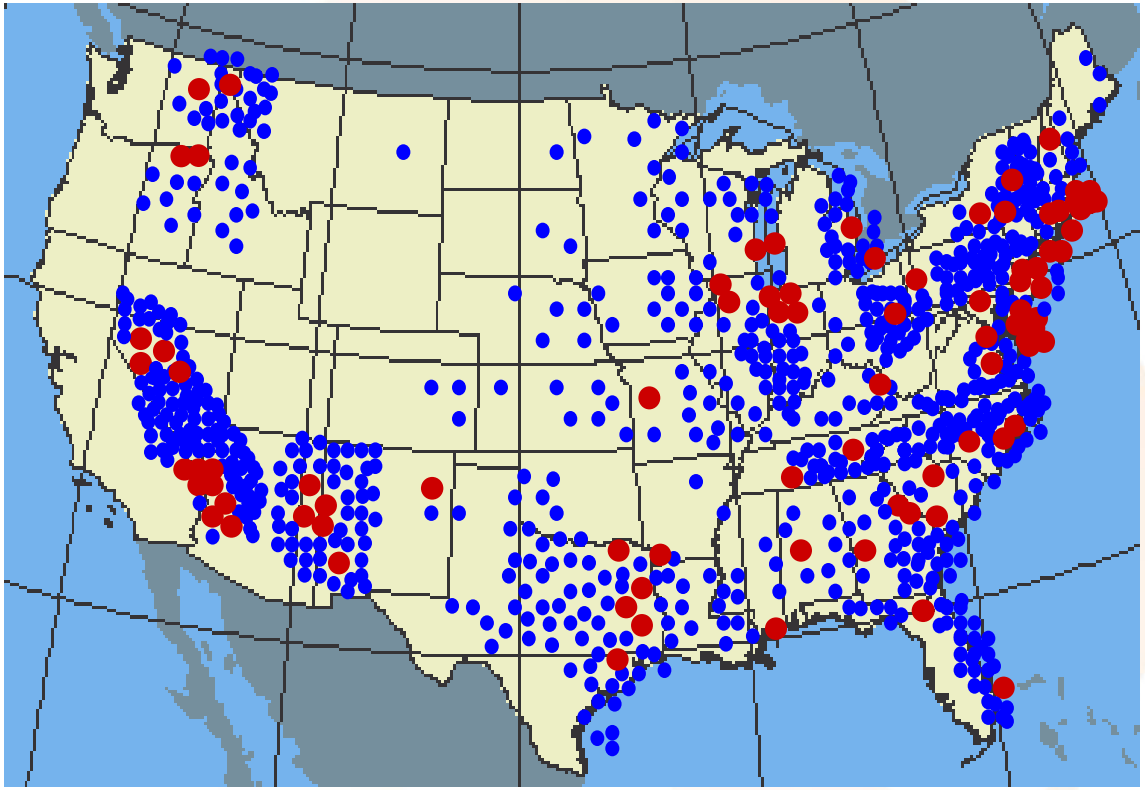
Results: We have developed a novel SNP genotype calling program, SNiPer-High Density (SNiPer-HD), for highly accurate genotype calling across hundreds of thousands of SNPs. The program employs an expectation-maximization (EM) algorithm with parameter

ing density is the long-held desire of geneticists to complete genome-wide association (GWA) studies using hundreds of thousands of SNPs to holistically scour the genome and identify associations between cases and controls which would allow for localization of functional mutations predisposing to disease (Craig and Stephan, 2005).. The exact number of SNPs to cover the majority of the genome is still being debated, but recent analysis of Phase II of the HapMap suggests that at least 250,000 well placed SNPs will be sufficient in a population with Asian or Caucasian descent (Altshuler, et al., 2005; Thorisson, et al., 2005). Both Affymetrix and Illumina now provide platforms allowing for genotyping of

Scientific Founders Perform WGA Studies Nationally

NIH Neuroscience Microarray Consortium

Total Projects: **455** (45,400 arrays)



- **1650** registered users
- **455** proposals submitted from about **114** institutions around the country and from over **287** different investigators

Founders' Technology and Experience

- 10 years of experience with genotyping platforms
- >100,000 expression profiles run
- >100,000 SNP arrays run (10k, 100k, 500k, >1M)
- Data warehousing solution
- First Affymetrix "Genomics Collaborators" in 2000
- First Affymetrix "Center of Excellence" in 2001
- First Affymetrix"TransMed" site in 2004
- NHLBI Programs in Genomic Applications (PGA)
- NEI intramural contract site
- NCI funded ALL catalog
- NIA funded Alzheimer's disease catalog
- NIH Neuroscience Microarray Consortium
- Autism Genome Project (AGP) Genotyping Site
- Center for Cancer Nanotechnology Excellence
- NCI funded Biomarkers Program
- FIND Consortium Genotyping Site
- ADNI Genotyping Site
- GAIN Genotyping Site
- ENDGAME
- Genotyping technologies (Illumina, Affymetrix, Sequenom)
- Sequencing technologies (Solexa, ABI SOLID, 454)

We know the Technical Issues Involved in Genome Scanning

Board of Directors

- **David Agus, M.D. (Co-founder)**: Research Director, Louis Warschaw Prostate Cancer CenterDirector, Spielberg Family Center for Applied Proteomics
- **Mari Baker**: President and CEO,Navigenics
- **David Brailer, M.D., Ph.D.:** Chairman Health Evolution Partners
- **John Doerr:** Partner, Kleiner Perkins Caufield & Byers
- **Mark Kvamme:** Partner, Sequoia Capital
- **Dana G. Mead, Jr.:** Partner, Kleiner Perkins Caufield & Byers
- **Sue Siegel:** Partner Mohr Davidow Ventures
- **Dietrich Stephan, Ph.D. (Co-founder):** Co-founder and Chief Science Officer, Navigenics; Deputy Director Emeritus, Translational Genomics Research Institute (TGen)

Clinical Advisory Board

- **David Agus, M.D., Chair**: Navigenics Co-founder, Research Director of the Louis Warschaw Prostate Cancer Center and Director of the Spielberg Family Center for Applied Proteomics at Cedars-Sinai Medical Center.
- **Carlos Camargo, M.D.**: Stanford Emeritus Clinical Professor of Medicine
- **Daniel Federman, M.D.:** Senior Dean for Alumni Relations and Clinical Teaching, Harvard Medical School
- **Bruce Landon, M.D., MBA**: Associate Professor of health care policy at Harvard Medical School; Associate Professor of Medicine at the Beth Israel Deaconess Medical Center. Practices internal medicine at BIDMC
- **Michael Nierenberg M.D.,** Medical Director, Navigenics, Emeritus Clinical Professor of Medicine, Stanford University

Scientific Advisory Board

- **Jeff Trent, Ph.D., Chair:** President and Scientific Director, Translational Genomics Research Institute (TGen)
- **David Botstein, Ph.D.**, Anthony B. Evnin Professor of Genomics, Director, Lewis-Sigler Institute for Integrative Genomics, Princeton University; Member, National Academy of Science
- **Isaac Kohane, M.D., Ph.D**: Henderson Associate Professor of Pediatrics and Health Sciences and Health Sciences and Technology, Harvard Medical School
- **Nicholas J. Schork, Ph.D.**: Director of Research, Scripps Genomic Medicine and Professor, Molecular and Experimental Medicine, The Scripps Research Institute
- **Dietrich Stephan, Ph.D.**: Co-Founder and Chief Science Officer, Navigenics, Deputy Director for Discovery Research, Translational Genomics Research Institute; Director and Senior Investigator, Neurogenomics Division, TGen; Chairman, NIH Neuroscience Microarray Consortium
- **Spencer Wells, Ph.D.**: director of Genographic Project, and National Geographic Explorer-in-Residence

Genetic Counseling Task Force

- **Robin Bennett. M.S., CGC, Chair**: Senior genetic counselor and clinical assistant professor at University of Washington
-
- **Peggy Conrad, M.S., C.G.C.,** Ms. Conrad is a practicing genetic counselor who helped to establish the University of San Francisco Familial Gastrointestinal Cancer Program. She has more than 10 years' experience in both the clinical and research aspects of hereditary cancer genetics.
- **Elissa Levin, M.S., C.G.C.** Genetics Counseling Program Director, Navigenics
- **Kelly Ormond, M.S., C.G.C.,** Associate Professor in the Department of Genetics and Director of the Human Genetics and Genetic Counseling graduate training program at Stanford University; past president, National Society of Genetic Counselors

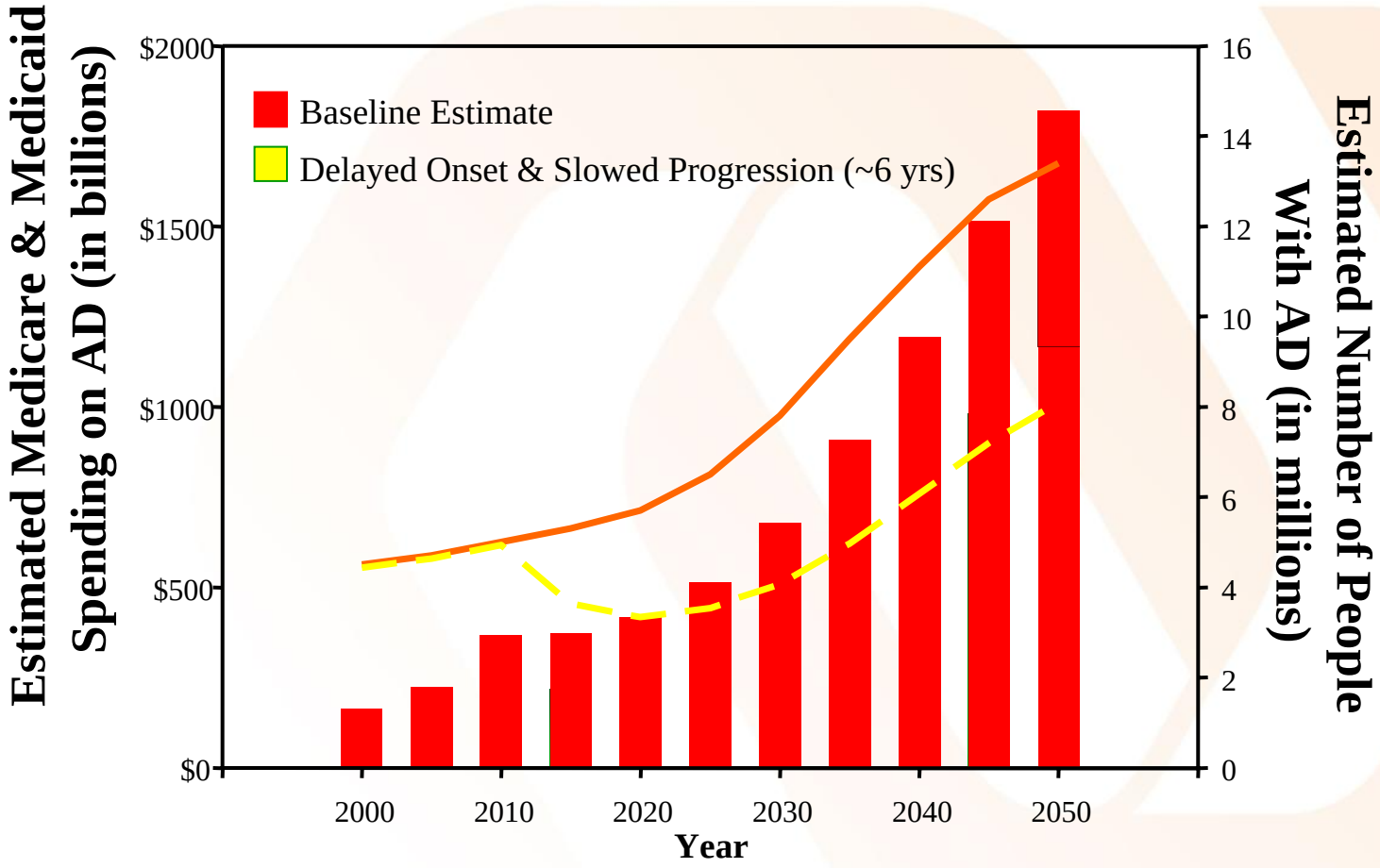
Policy and Ethics Task Force

- **Greg Simon, J.D., Chair,** President of Faster Cures/the Center for Accelerating Medical Solutions, which helps accelerate the discovery, development and deployment of new medical treatments
- **Rachel Grob, M.A., Ph.D.,** Associate Dean of Graduate Studies, Director Child Development Institute, Sarah Lawrence College, and is also an Investigator in Health Policy Research, funded by the Robert Wood Johnson Foundation, 2006-09
- **Kathi Hanna, M.S., Ph.D.,** analyst, writer, and editor specializing in biomedical research policy, previously, Research Director and Editorial Consultant to President Clinton's National Bioethics Advisory Commission
- **Paul Slovic, Ph.D.,** Founder and President, Decision Research, a non-profit research organization investigating human judgment, decision-making, and risk

The Looming Health Care Crisis

Navigenics

Estimated Savings in Prevalence & Costs of AD with Delayed Onset/Progression



Adapted from The Lewin Group Report, June 2004, "Saving Lives. Saving Money: Dividends for Americans Investing in Alzheimer Research," The Alzheimer's Association (http://www.alz.org/Resources/FactSheets/Lewin_FullReport1.pdf)

Other Facts

- Estimated that Medicaid will be depleted in ~10 years at current trajectory
- Estimated that we will be paying in excess of 40% of our National GDP to healthcare to treat chronic disease
- An aging population, and a growing population, shows similar growth curves for diabetes, heart disease, etc

- We must implement effective prevention strategies NOW, and we believe that genetic risk information can assist with that goal.

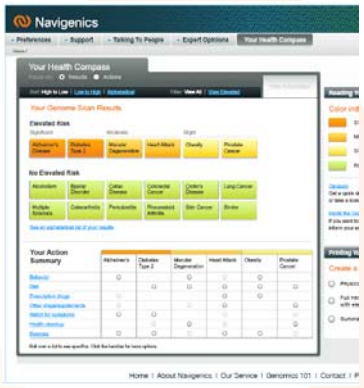
The Navigenics Sreen (common and rare variants)

Navigenics

5-Step Service Offering

1	2	3	4	5
Customer Acquisition	Laboratory	Bioinformatics	Personalized Web Portal	Ongoing Service


```
ATACCGCTGGCCCTT
TGGCATTACCTATGA
AGATTGCTTCAGCCA
GCGTCAGTTTCAACC
TGTACGCTAGTGTGT
TTCTACTCACGTGTC
TCAGCATTGATCGAT
ACCTGGCTATTGTTC
ACCCAATGAAGTCCC
```



nature genetics

The NEW ENGLAND JOURNAL of MEDICINE

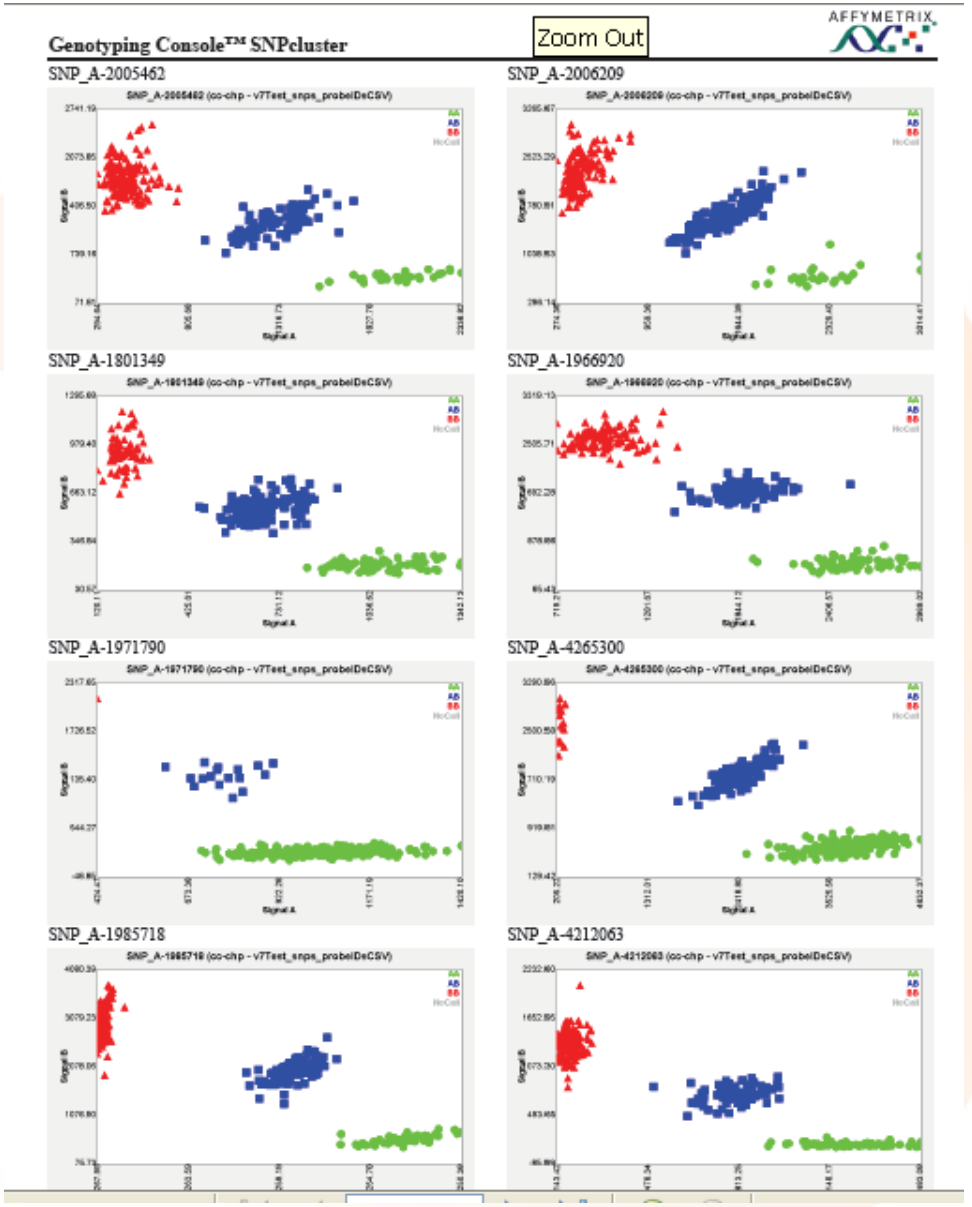
FUTURE: Full genome sequencing, copy number analysis, methylation status leading to personalized exposure mitigation strategies and biomarker monitoring programs fully integrated into the established health care system.

Stringent Curation Criteria for Common Variants

- **Replication in the same ethnic group**
  - Once for GWAS, twice for candidate gene studies
  - >60% independent sample sets show same statistically significant effect with same allele (after trimming underpowered samples)
- **Study design** - An effort was made to sample controls from the same source population as the cases, e.g. ethnicity, gender, age, or other risk factors.
- **Reasonable sample size to detect weak effects**. OR <1.5 needs 250 cases/250 controls at least.
- **Significance level** - Exact value depends on magnitude of the study (e.g. GWAS or candidate gene)
  - **Sound statistical design** - correction for multiple testing, population stratification, confounding
  - **Sound laboratory practice** - independent genotyping platforms, replicated samples
  - **Functional data** and **magnitude of effect** are also taken into account, but studies are not automatically excluded if functional data is unavailable or the effect estimate is small.

QUALITY

CLIA and stringent QC lab
Captured perfectly
Per SNP algorithm checks
Per SNP concordance
H-W equilibrium checks

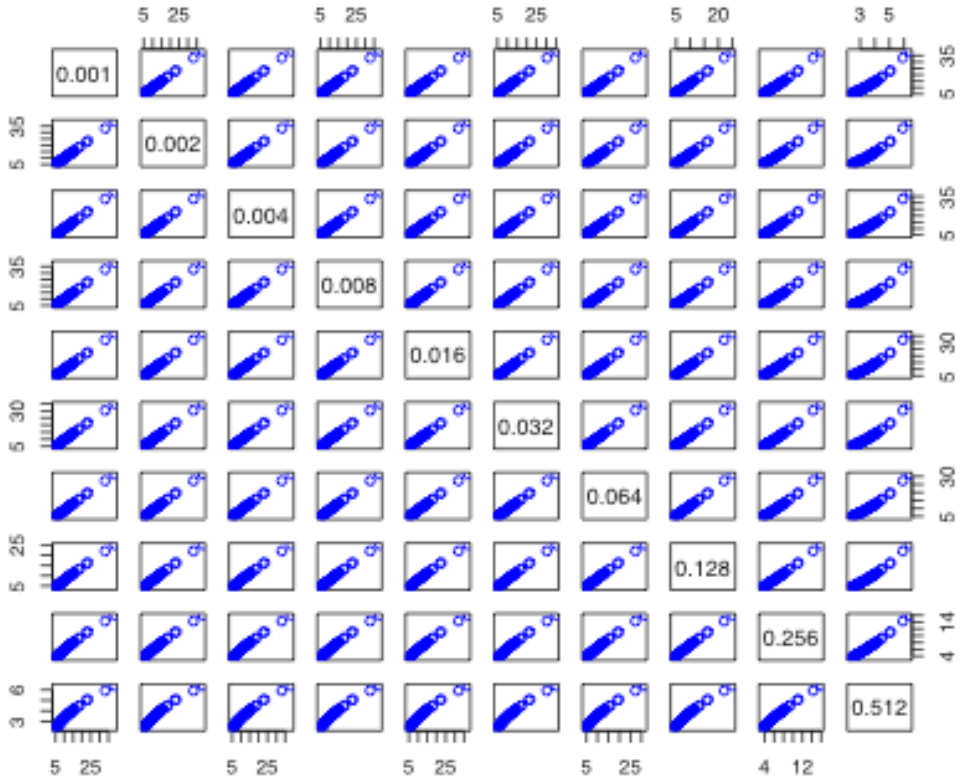


Genetic Component Index – see full details at navigenics.com

- A model-based risk estimate
- A higher GCI score can be intuitively interpreted as an increased risk for a condition:
  - We assume that the odds ratio values are available
  - We assume that the prevalence of the condition is known
  - We assume that the genotype frequencies in the proper population are known
  - We assume that the test individuals are from the same ancestry background as the populations used for the studies and as the HapMap/reference pop
  - We assume that the amalgamated risk is a product of the different risk alleles of the individual SNPs (acting independently)

Effect of Prevalence on GCI for Type 2 Diabetes – see full details at navigenics.com

- Based on HapMap CEU data
- The rank of an individual does not vary as a function of the frequency.

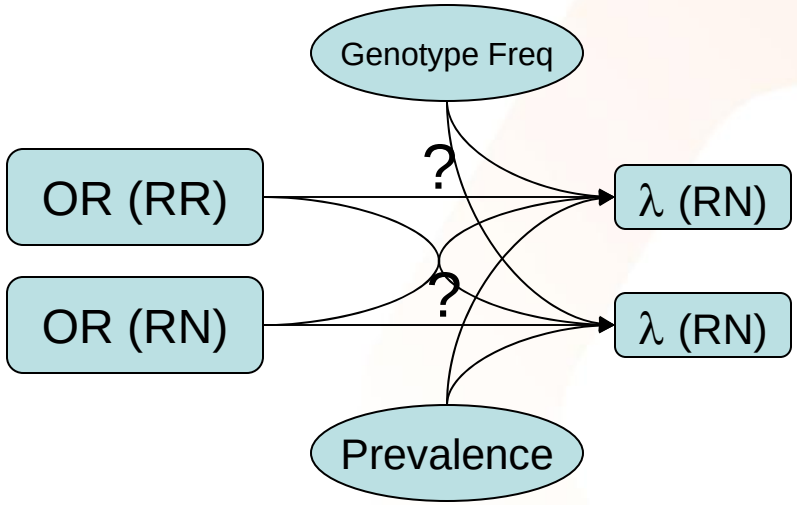


- The absolute values varies:

$$P(D \mid g_i) = \lambda_i \cdot \frac{p}{\sum_{i=0}^{k} f_i \lambda_i}$$

Output: “Your genetic load for T2DM exceeds 95% of the population”

Finding the Relative Risk - see full details at navigenics.com



$$C = \sum_{i=0}^{k} f_i \lambda_i$$

$$1 = \sum_{i=0}^{k} \frac{OR_i f_i}{C - p + OR_i p}$$

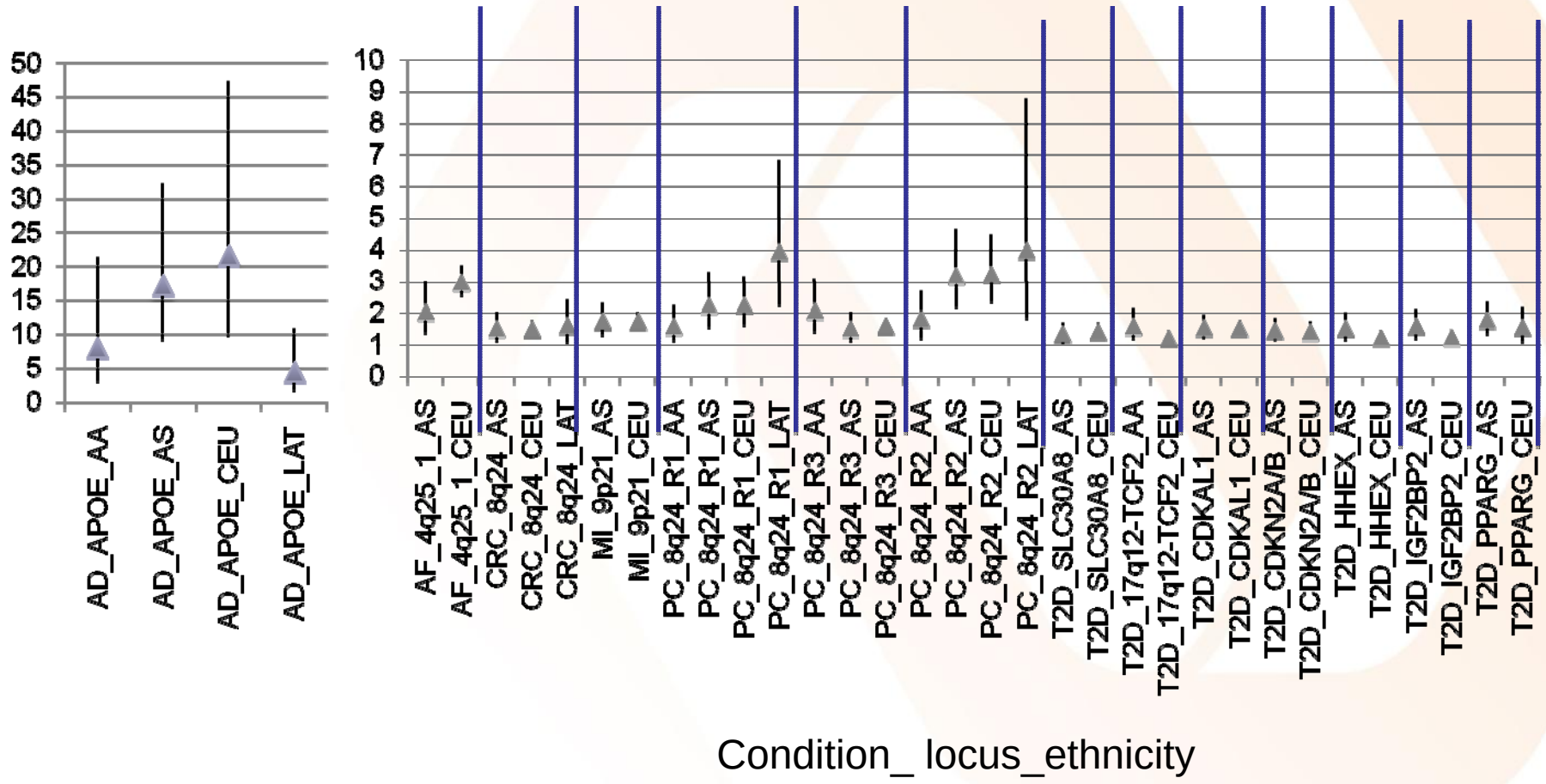
- We normally get genotypic odds ratios RR/NN, RN/NN
- Using genotype frequencies and prevalence, we derive a set of quadratic equations – the solution provides the relative risks.

Assumptions in the Estimated Absolute Risk Measure – see full details at navigenics.com

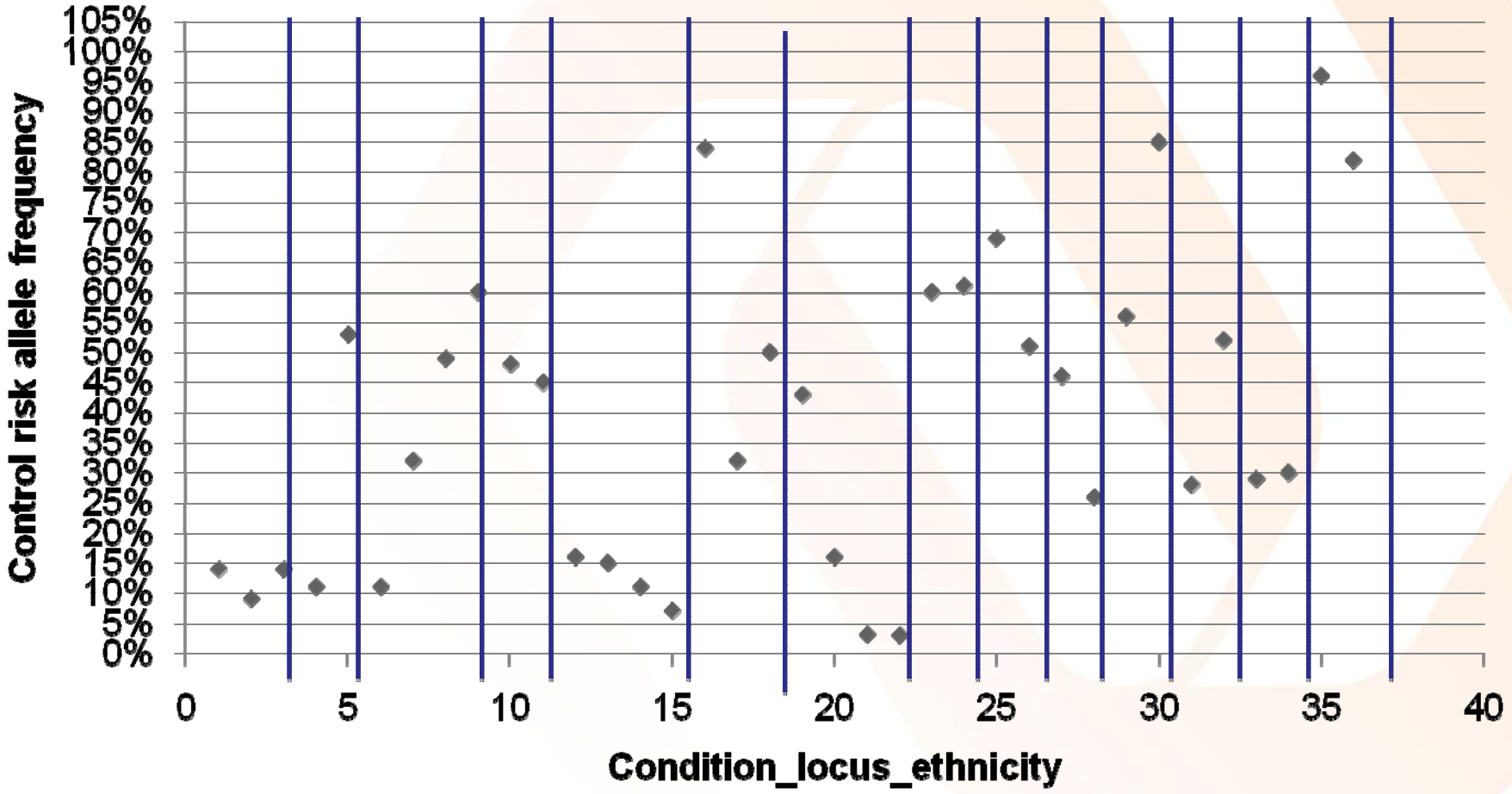
- A visualization tool to impart probabilistic risk assessment to an individual based on their combinations of all known genetic risk factors

- All of the GCI assumptions
- The lifetime risk, age-defined prevalence, or age-defined incidence of the condition is known
- The lifetime risk for an individual is proportional to the individual's GCI score divided by the average GCI score in 60 Caucasian individuals

Odds-ratios for different ethnicities are usually similar



Risk allele frequencies in controls for different ethnicities are usually different



Capturing Genotypes and Automated Error Correction

ARTICLE

Leveraging the HapMap Correlation Structure in Association Studies

Noah Zaitlen, Hyun Min Kang, Eleazar Eskin, and Eran Halperin

Recent high-throughput genotyping technologies, such as the Affymetrix 500k array and the Illumina HumanHap 550 beadchip, have driven down the costs of association studies and have enabled the measurement of single-nucleotide polymorphism (SNP) allele frequency differences between case and control populations on a genomewide scale. A key aspect in the efficiency of association studies is the notion of "indirect association," where only a subset of SNPs are collected to serve as proxies for the uncollected SNPs, taking advantage of the correlation structure between SNPs. Recently, a new class of methods for indirect association, multimarker methods, has been proposed. Although the multimarker methods are a considerable advancement, current methods do not fully take advantage of the correlation structure between SNPs and their multimarker proxies. In this article, we propose a novel multimarker indirect-association method, WHAP, that is based on a weighted sum of the haplotype frequency differences. In contrast to traditional indirect-association methods, we show analytically that there is a considerable gain in power achieved by our method compared with both single-marker and multimarker tests, as well as traditional haplotype-based tests. Our results are supported by empirical evaluation across the HapMap reference panel data sets, and a software implementation for the Affymetrix 500k and Illumina HumanHap 550 chips is available for download.

Large-scale case-control association studies are a potentially powerful tool for discovering the genetic basis of human disease.[1–3] Recent high-throughput genotyping technologies, such as the Affymetrix 500k array and the Illumina HumanHap 550 beadchip, have driven down the costs of association studies and have allowed us to measure allele frequency differences between case and control populations on a genomewide scale.[4,5] A key aspect in the efficiency of association studies is the notion of "indirect association." By leveraging the linkage disequilibrium (LD) structure of the genome, frequency differences between case and control populations do not need to be measured in all SNPs but only in a subset, or a set of tag SNPs that serve as proxies for the remaining uncollected SNPs (we also refer to the uncollected SNPs as "hidden SNPs").[6] A chromosome carrying a particular allele of a tag SNP has a high probability of carrying a particular allele of a proximal hidden SNP. Thus, an allele frequency difference in a hidden SNP will manifest itself as an allele frequency difference in a tag SNP. This correlation is often measured between two SNPs by the correlation coefficient r^2. The r^2 measure is widely used in the design and analysis of association studies, because the relation between the power of detecting an association at the hidden SNP and only observing the tag SNP has been well understood for some time (e.g., see the work of Pritchard and Preworzski[7] and Sham et al.[8]).

Tag SNPs are chosen by examining the LD structure of a reference panel such as the HapMap,[9] which is a data set that contains a complete set of genotypes from 270 individuals, with >3.9 million SNPs across the genome. Choosing a set of tag SNPs is a challenging problem, since the LD structure is quite complex and varies through the genome. To date, many tag SNP selection methods have been proposed.[10,11] These methods employ different statistical criteria, the most common being procurement of a set of tag SNPs for which every hidden SNP is "covered" by a tag SNP, such that the correlation coefficient r^2 between the two SNPs in the reference set is higher than a certain threshold (e.g., see the work of Carlson et al.[11]). These methods vary greatly in the optimization methods used to obtain the tag SNPs.

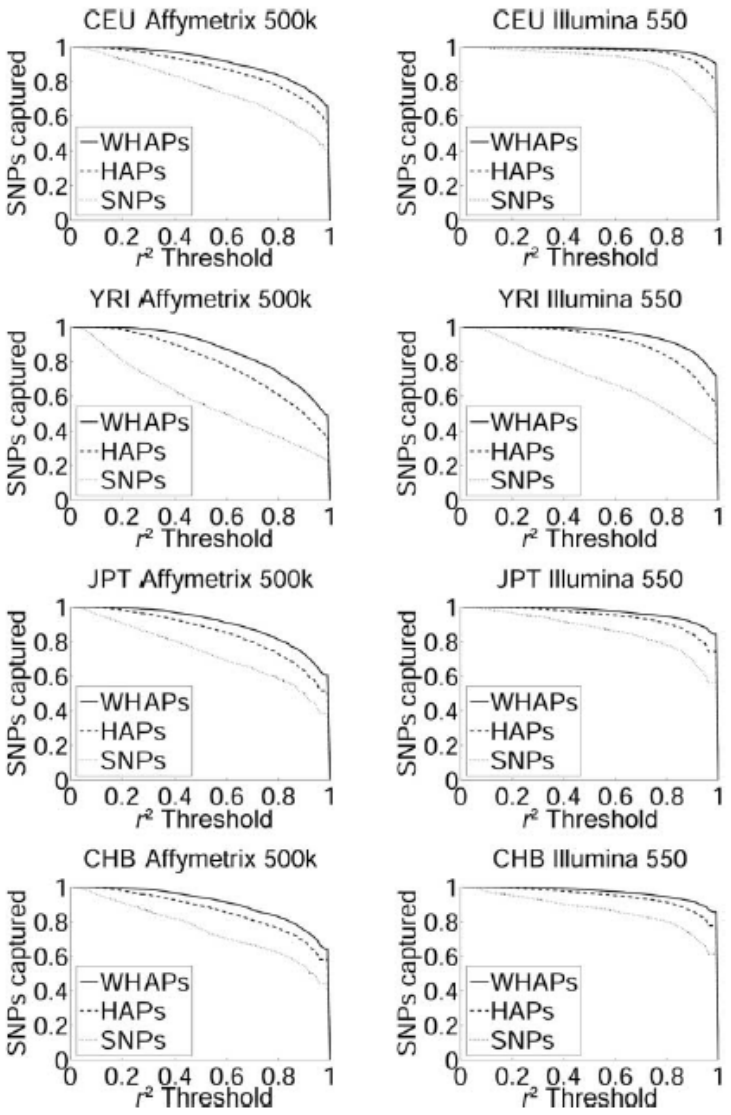
Recently, a new class of methods—multimarker methods—has been proposed.[10,12–14] These methods take advantage of the fact that some pairs (or groups) of SNPs serve as better proxies for the hidden SNPs than does any single SNP. Since multimarker proxies have more than two possible alleles, the frequencies of a specific sequence of alleles in these SNPs (i.e., a haplotype) are compared between the cases and the controls. Thus, a specific haplotype, instead of a single SNP, is used as a proxy for a hidden SNP. It has been shown empirically that these methods can reduce the number of tags required to achieve equivalent power.[10] In addition, it has been empirically shown that even if the set of tag SNPs is fixed—such as in the case where a commercial high-throughput genotype prod-

From the Bioinformatics Program (N.Z.) and Department of Computer Science and Engineering (H.M.K.), University of California–San Diego, La Jolla, CA; Departments of Computer Science and Human Genetics, University of California–Los Angeles, Los Angeles (E.E.); and International Computer Science Institute, Berkeley, CA (E.H.)

Received November 2, 2006; accepted for publication January 24, 2007; electronically published March 2, 2007.

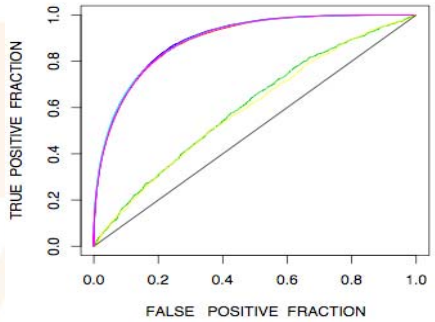
Address for correspondence and reprints: Dr. Eran Halperin, International Computer Science Institute, 1947 Center Street, Suite 600, Berkeley, CA 94704. E-mail: heran@icsi.berkeley.edu

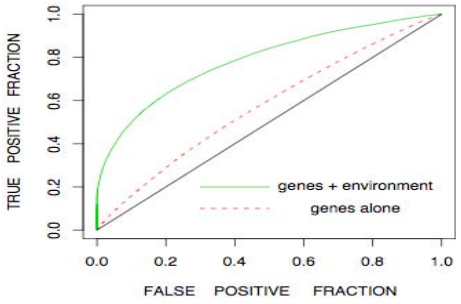
Am. J. Hum. Genet. 2007;80:683–691. 0002-9297/2007/8004-0010$15.00
DOI: 10.1086/513109

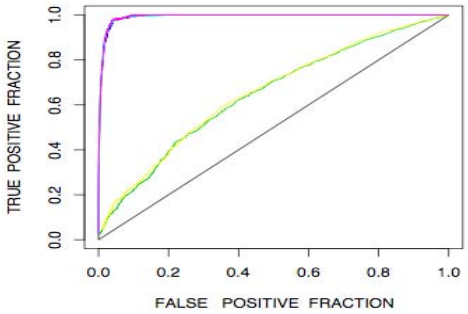


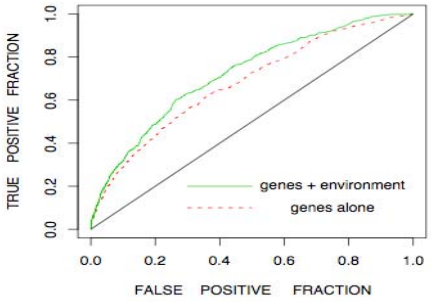
Estimated Genetic Variance we have Today

- Large effect sizes have been found
- No GxE



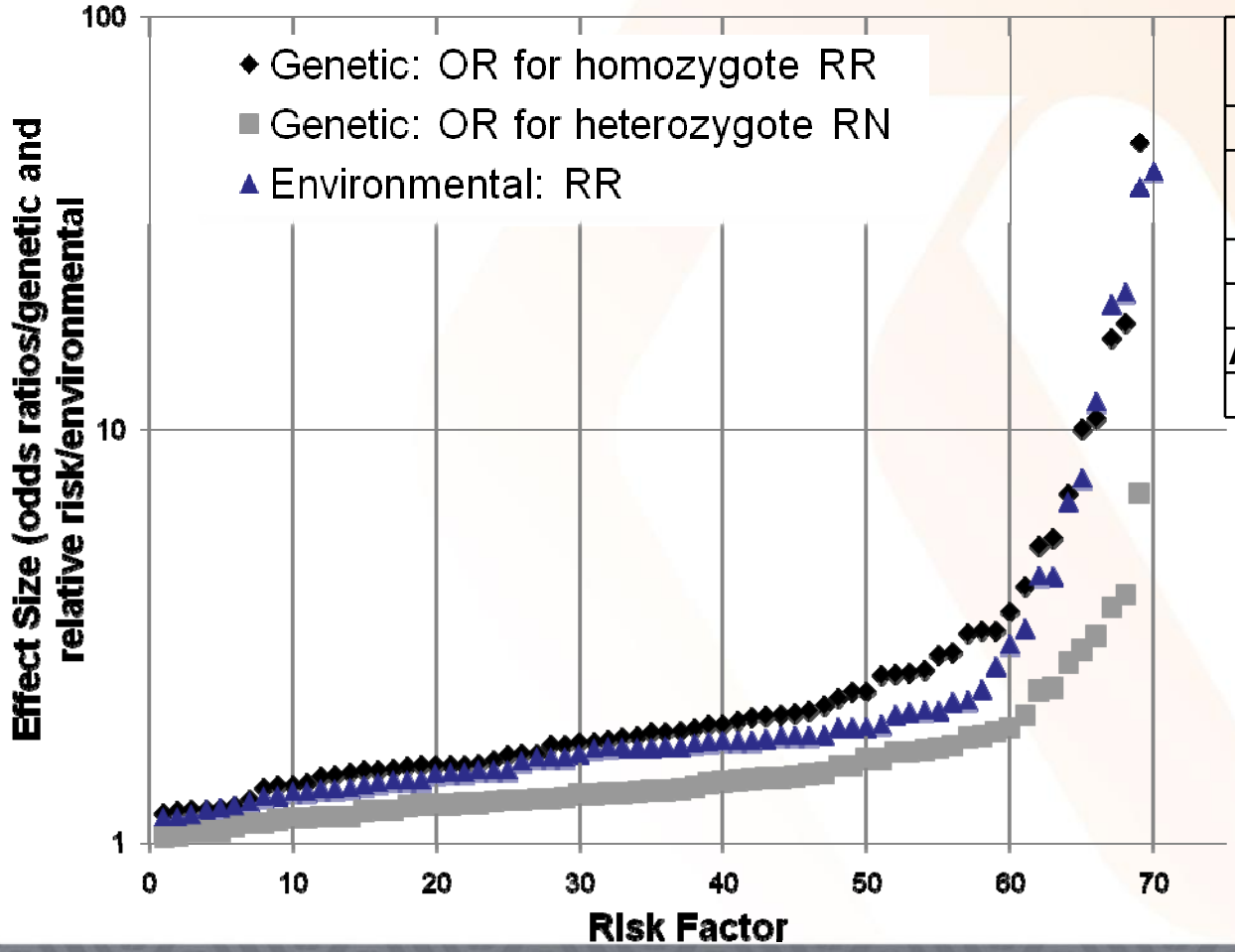






Disease	Relative risk of homozygous risk	Relative risk of heterozygous	Estimated number of unknown variants	Fraction of genetic variation explained by known variants out of the entire GENETIC variation
Type 2 Diabetes	1.10	1.05	1600	7%
Crohn's Disease	1.10	1.05	13958	4.4%
Rheumatoid Arthritis	1.10	1.05	6237	14.4%

Effect Sizes for "Genetic Risk Factors" vs. Environmental Risk Factors are Similar



risk factor	*condition*	*effect size*
Ex-smoker	T2D	1.15
PPARG genotype	T2D	1.53
HDL<35mg/dl	CHD	2.08
MHC genotype	RA	5
APOE genotype	AD	18
BMI>35	T2D	42

Risk factors determined from literature using strict curation guidelines

State-of-the-art clinical risk assessment: MI

Grade 2-4 hypertension	1.92
LDL>160	1.74
HDL<35	1.46
Smoker (last 12 mo)	1.71
T2DM	1.47
No exercise	1.39

State-of-the-art Clinical Risk Assessment: Myocardial Infarction

Grade 2-4 hypertension	1.92
LDL>160	1.74
HDL<35	1.46
Smoker (last 12 mo)	1.71
T2DM	1.47
No exercise	1.39
9p21	1.72
MTHFD1L	1.53

Risk Communication

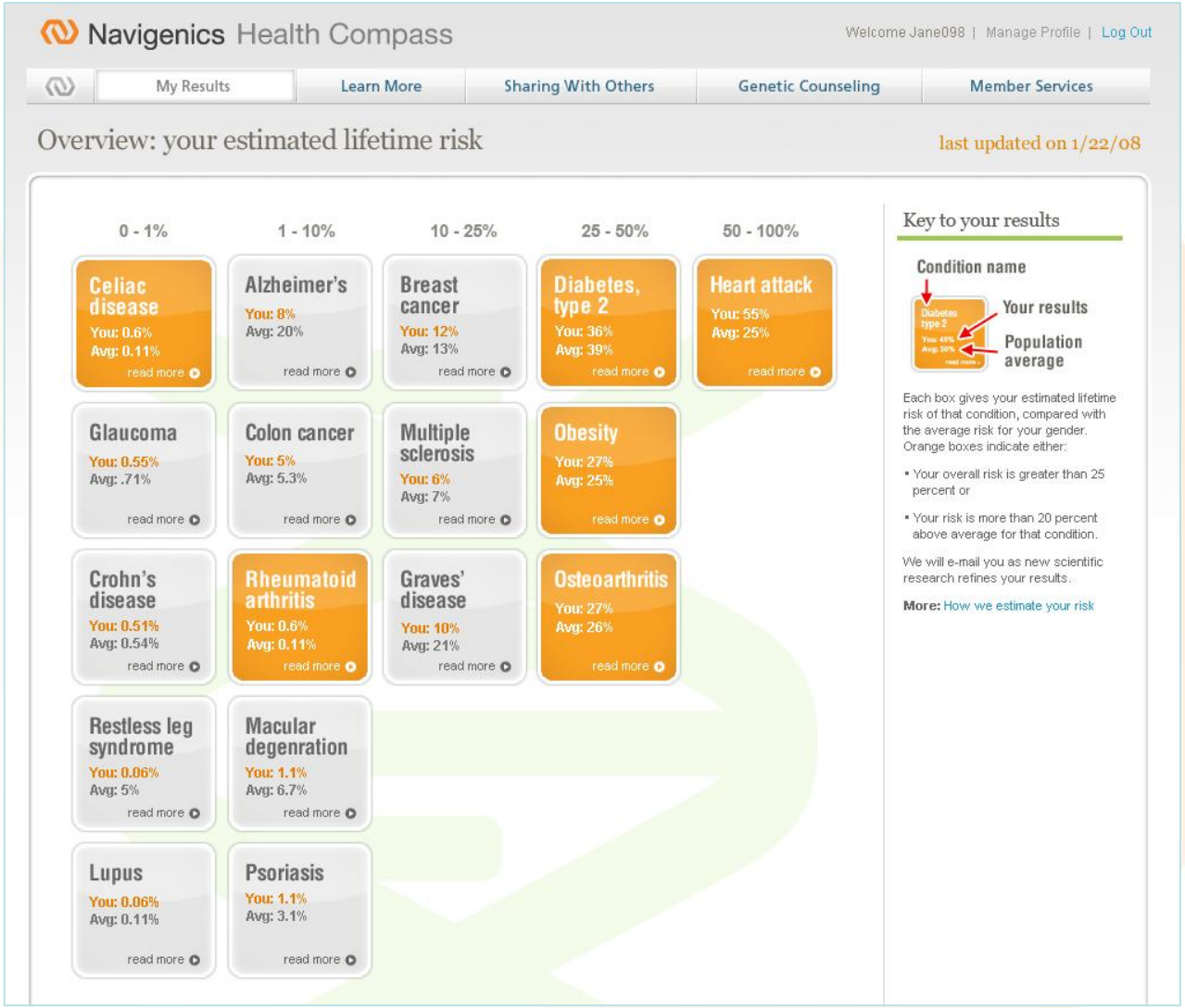
Navigenics

Professional Access

- Genetic counselors at any time included in the Navigenics service
- Tools to talk to your doctor
- Website was built with input from physicians, genetic counselors, medical journalists, risk communication experts to make it understandable for a non-expert individual.

Health Compass: Results Overview

People want to know what this means *for them*

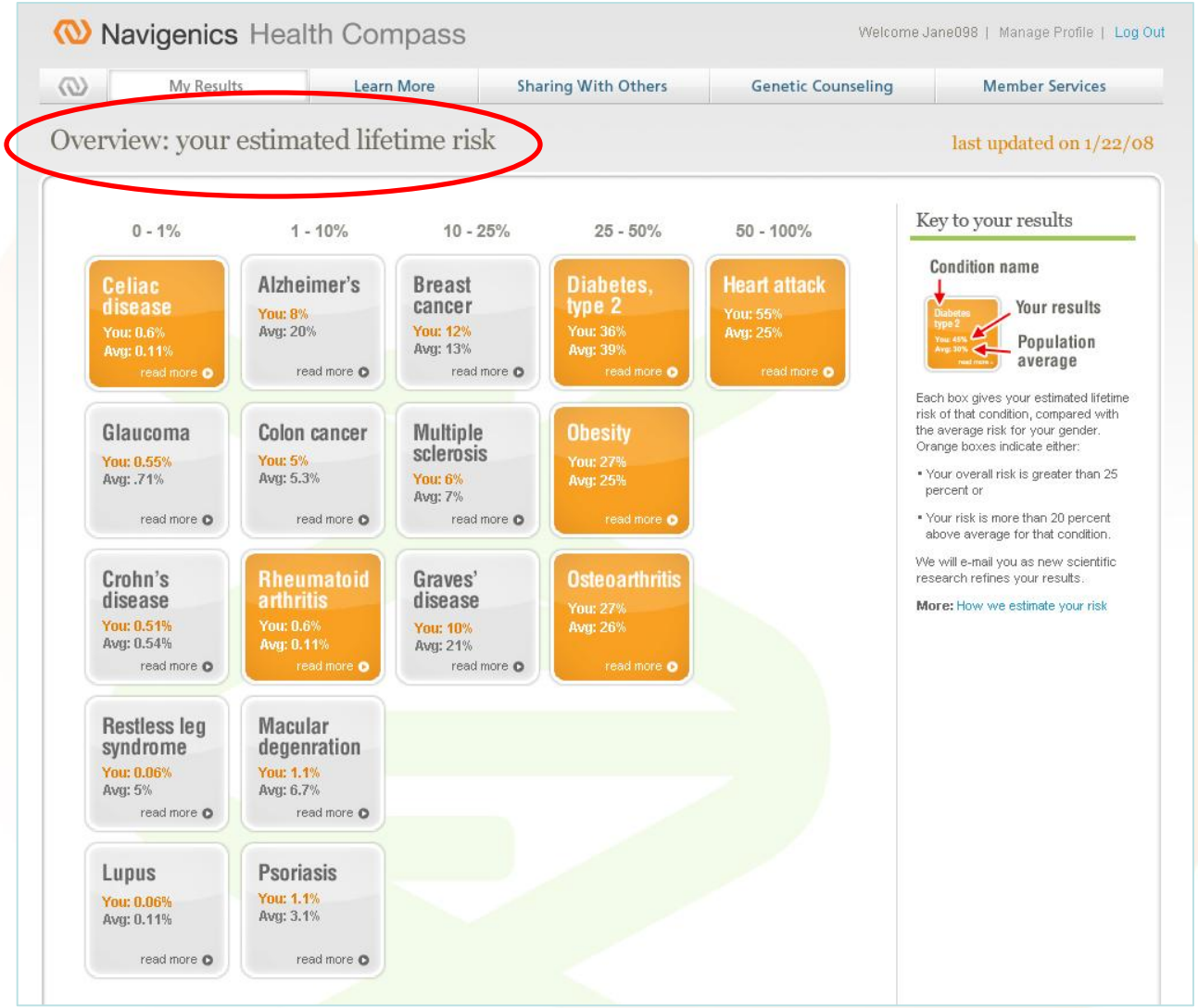


Health Compass: Results Overview

People want to know what this means *for them*

Estimated Lifetime Risk

Take the general population LTR and refine based on the individual's genotypes



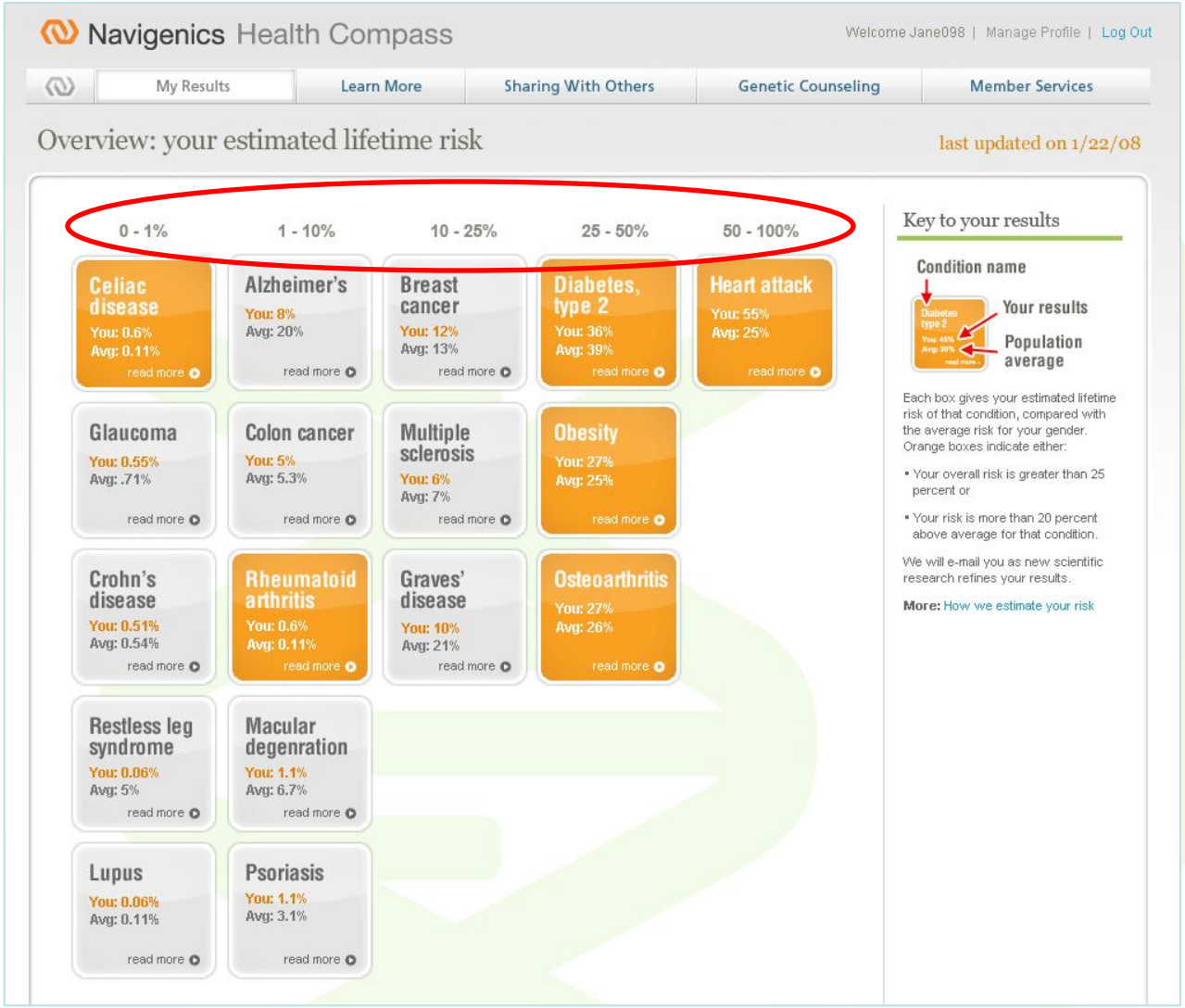
Health Compass: Results Overview

People want to know what this means *for them*

Estimated Lifetime Risk

Take the general population LTR and refine based on the individual's genotypes

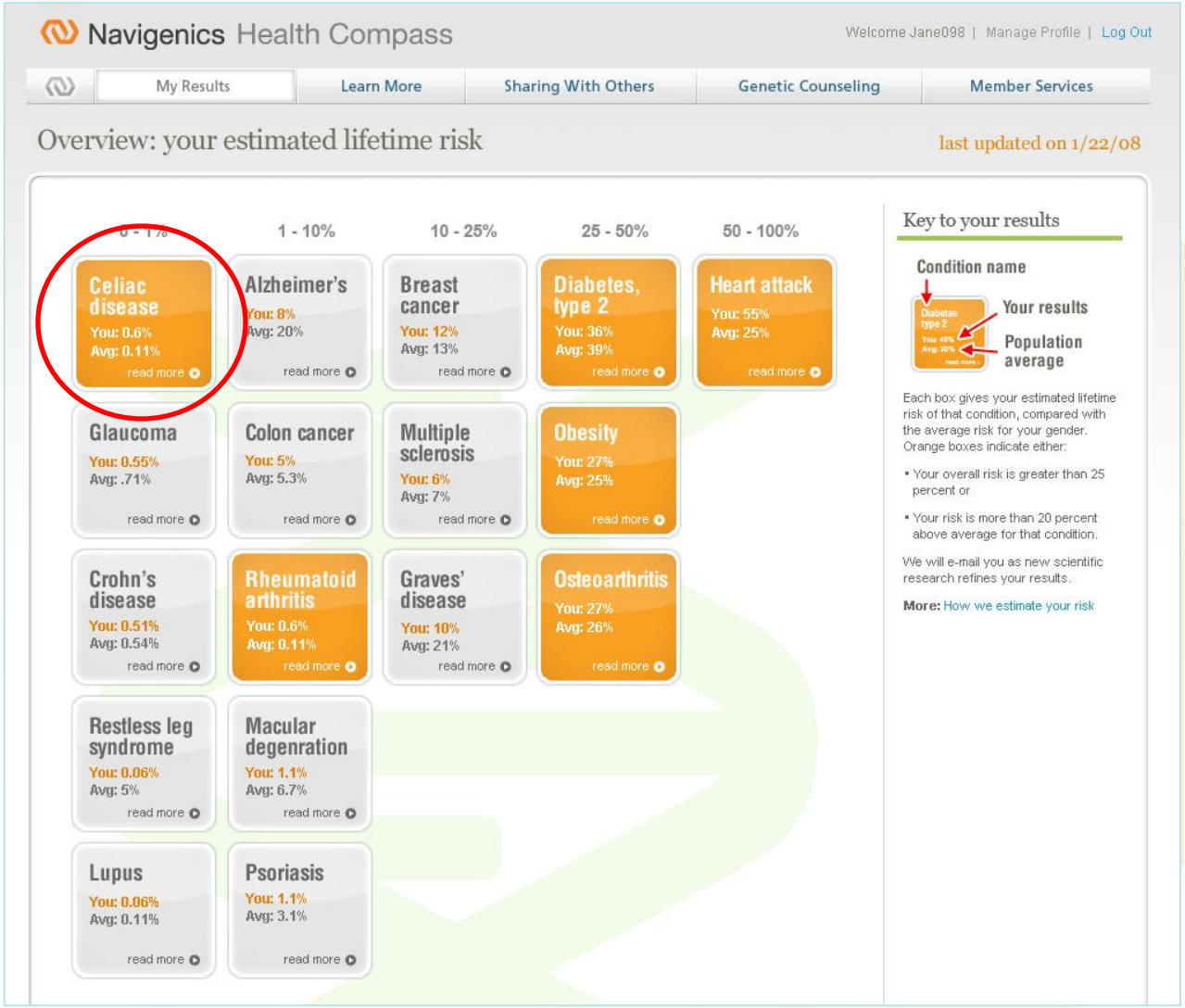
Place the conditions into "buckets" to highlight the overall LTR estimate



Health Compass: Results Overview

Orange Box

Estimated LTR is 20% or more than the general population

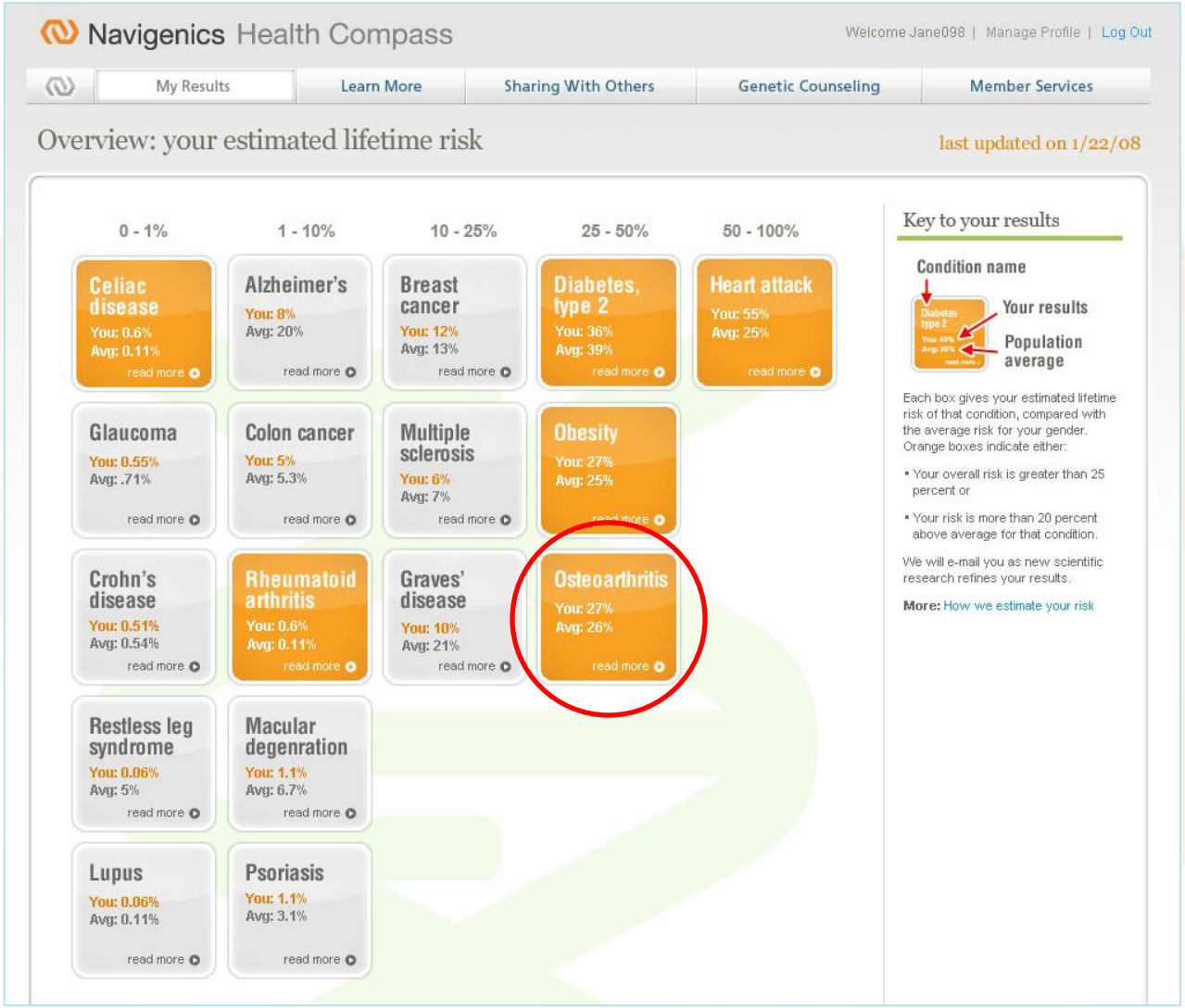


Health Compass: Results Overview

Orange Box

Estimated LTR is 20% or more than the general population

Estimated LTR is more than 25% total



Health Compass: Results Overview

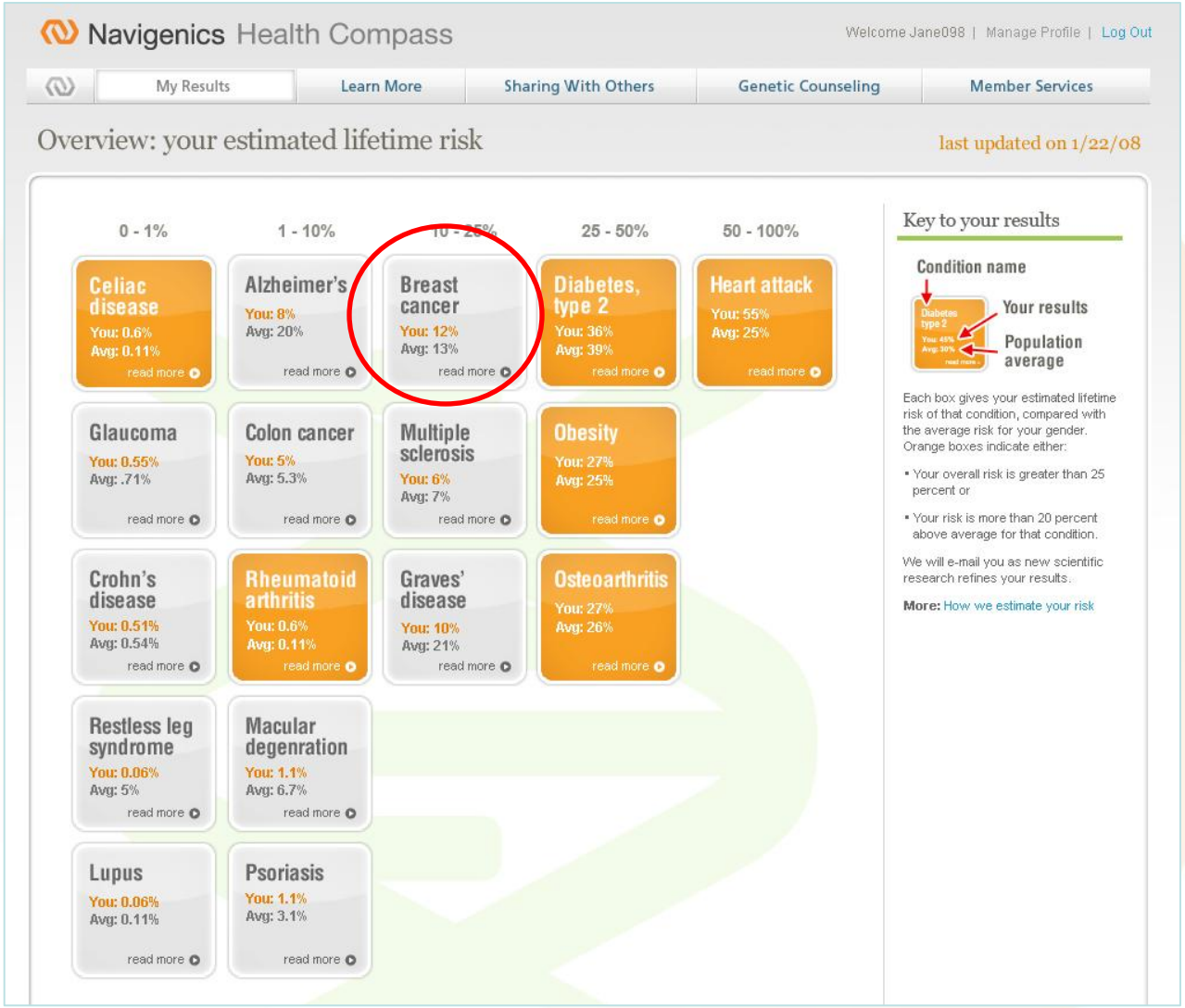
Orange Box

Estimated LTR is 20% or more than the general population

Estimated LTR is more than 25% total

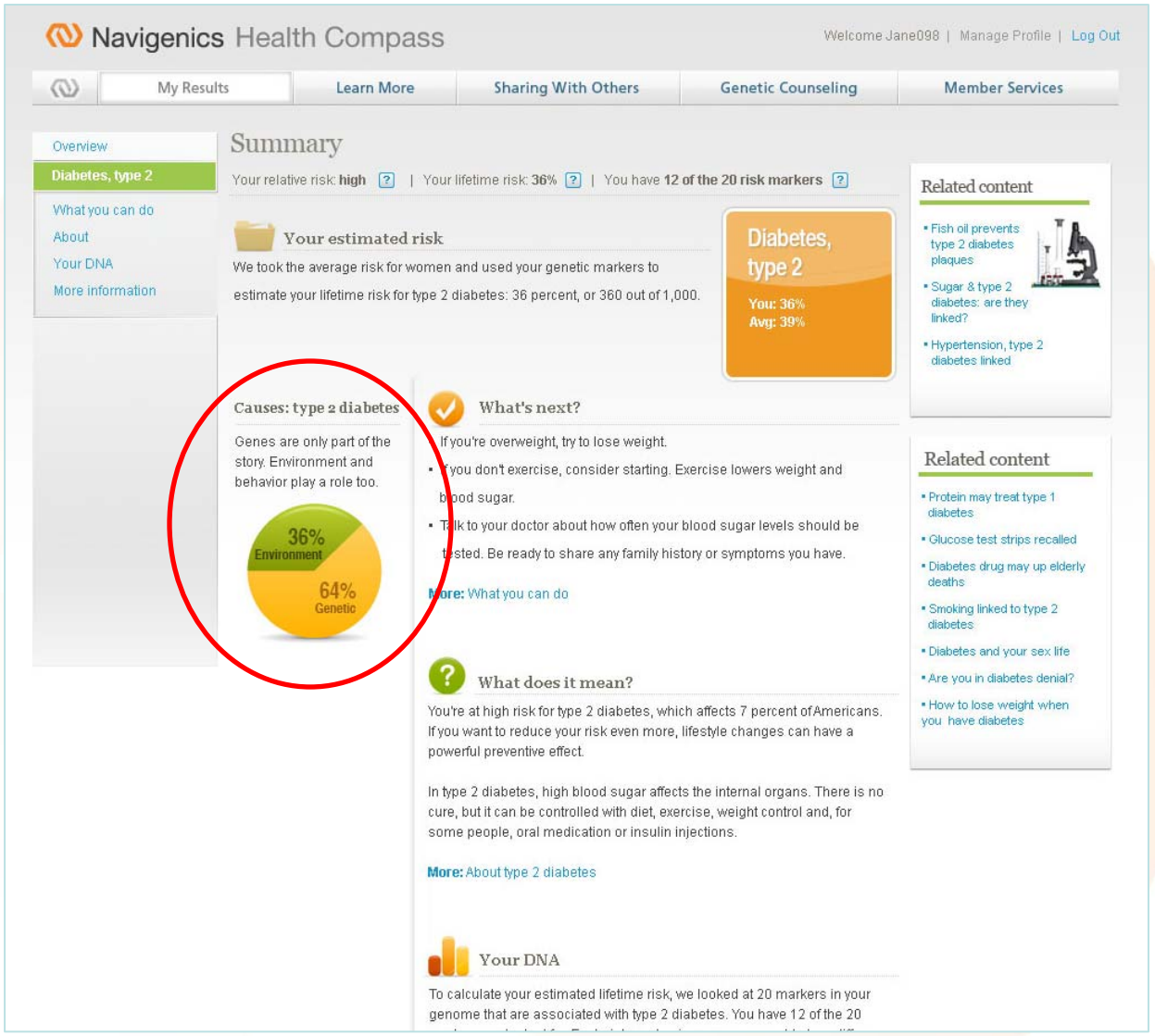
Gray Box

Estimated LTR is at or below the population average



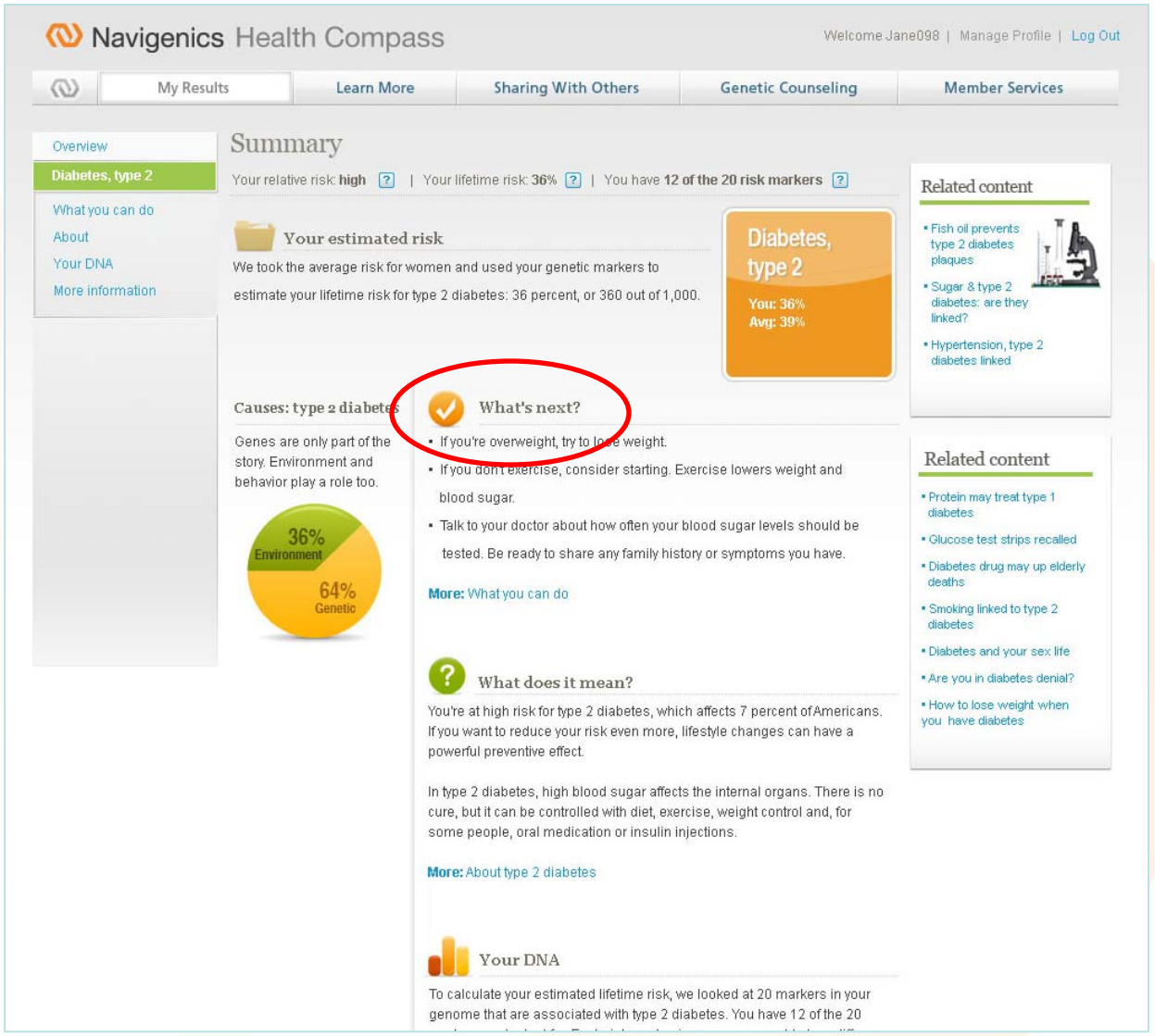
Condition-Specific Summary

- In-depth report for each condition
- Highlight genetic vs. environmental contribution to disease



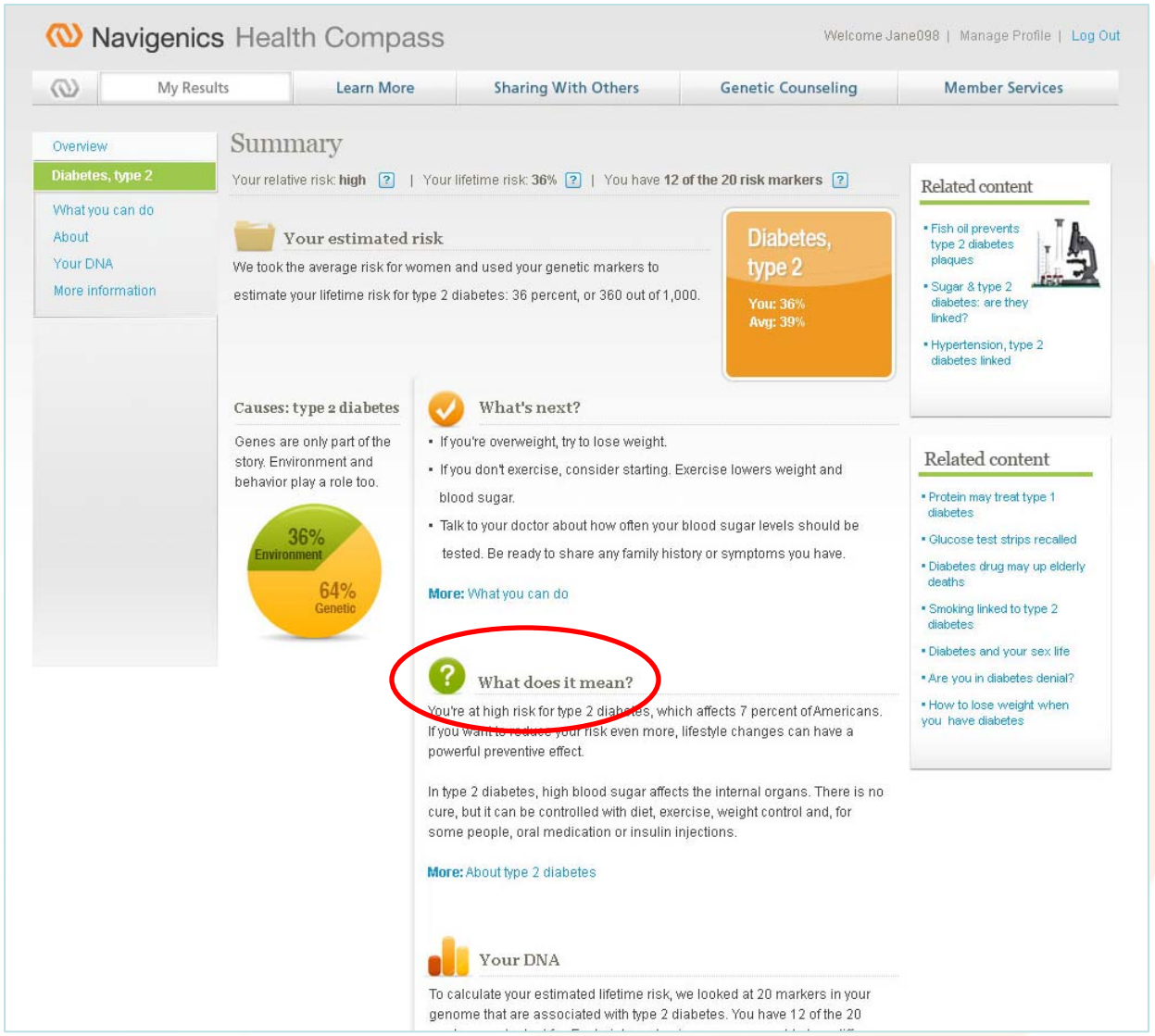
Condition-Specific Summary

- In-depth report for each condition
- What's next?



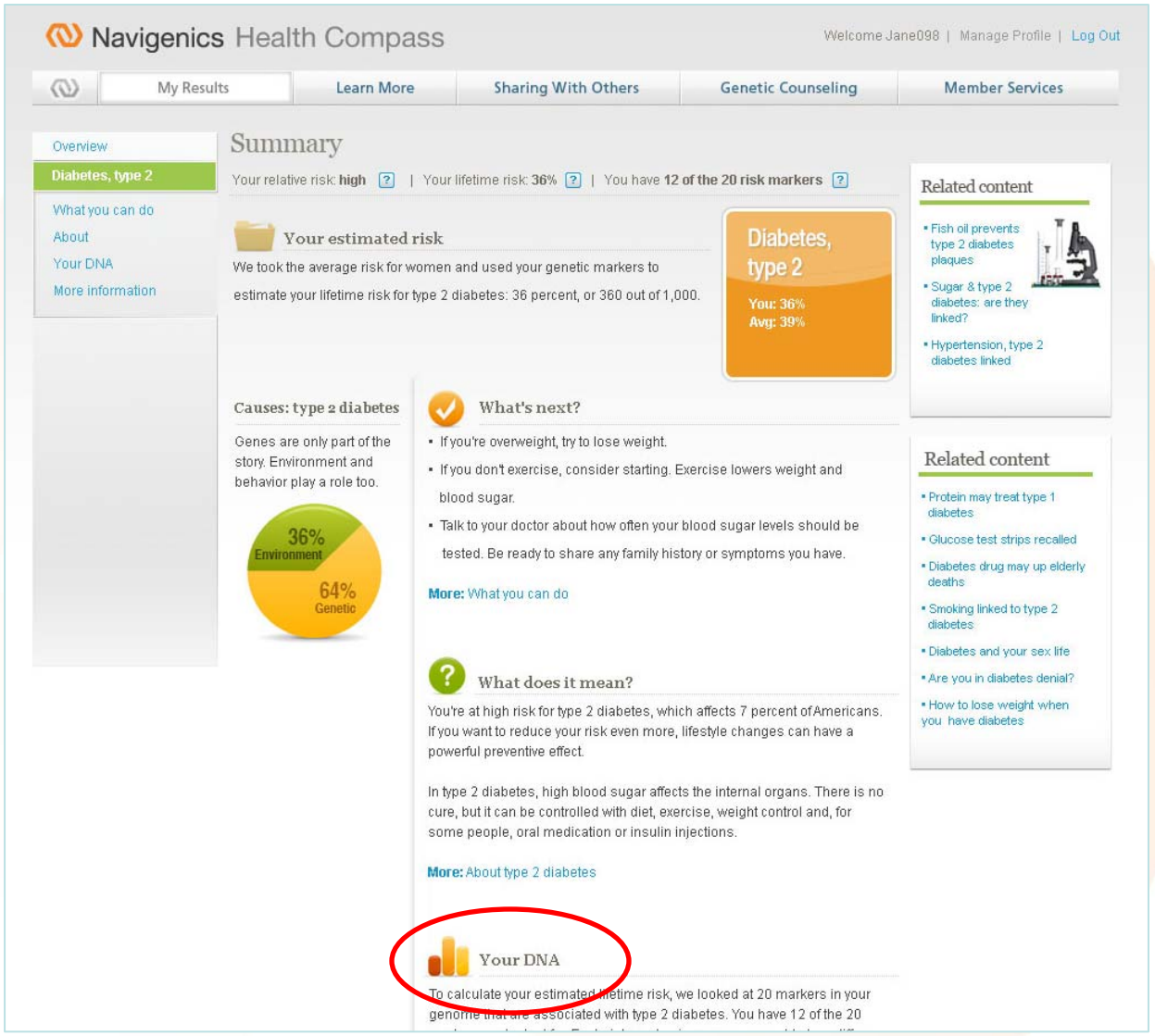
Condition-Specific Summary

- In-depth report for each condition
- What's next?
- What does it mean?

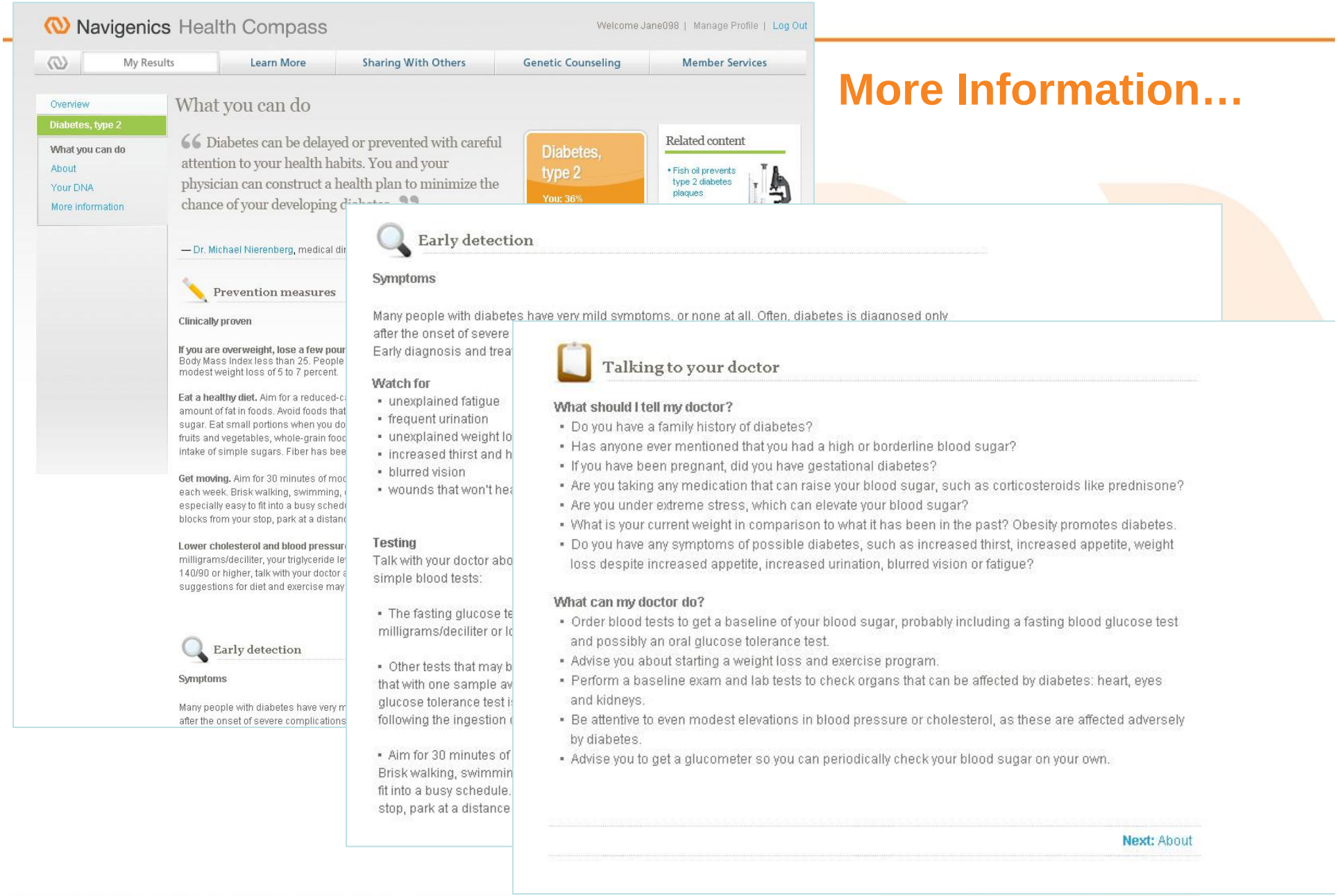


Condition-Specific Summary

- In-depth report for each condition
- What's next?
- What does it mean?
- Your DNA
- Total risk markers identified
- SNPs included in analysis
- Effect of genotype
- Primary resources



More Information…



Evidence of behavior change after genetic testing

Aspinwall. Cancer Epidemiology Biomarkers and Prevention 17:1510. 2008

- Melanoma screening considered controversial
  - Positive test results cause distress?
  - Negative test lead to false sense security?
  - No advantage over counseling based on family history alone?
- Prospective study of 59 individuals at high risk for melanoma
  - All had family history of melanoma
  - Tested for high penetrance CDKN2A/p16 mutation
- Improved screening behaviors in mutation carriers
  - Increased intention to obtain body skin exams
  - Increased intention and adherence to skin self exam
  - Increased number of body sites examined
- No evidence of false sense of security among non-carriers
  - Did not decrease screening behaviors
  - Improved adherence to skin self exam
- Genetic testing superior to counseling based on family history
  - Study participants had been twice notified of their high risk based on fam hx
  - Yet, poor baseline compliance with screening recommendations

How Navigenics Operates Differently (outside of scientific rigor)

Navigenics

Security / Privacy

- We operate in a HIPAA consistent manner
- We require opt-in for internal research and/or third party research
- Privacy and security policies ensure that our members can feel comfortable and confident receiving genetic information and analyses, and that they alone control how that information is to be used and distributed.
- We use the most advanced data protection systems; we safeguard, maintain and update your genetic profile in a highly secure environment. All customer profiles are anonymous to assure data security.
- Although there is concern about insurance companies misusing genetic information, there are currently no cases on record of this happening. We are very diligent about communicating how to avoid this problem to our members.

Physician Educational Initiatives:

- **TOP DOWN**: Ongoing education of the “physician’s physician” at leading clinical centers in the country such as the Mayo Clinic, Scripps, Harvard, Duke, and the Cleveland Clinic.
- **DIRECT ENGAGEMENT**: Navigenics sponsored Genomic Medicine CME training program with Medscape. This course generated >5,000 readers within the first two months, with 99.6% of readers completing the entire course
- **BOTTOM UP**: A physician portal to the Navigenics product is provided. This site explains additional scientific details that the physician can use to learn about the product and how it can help their patients.

The Great Debate

Common Arguments:

- Analytic validity – is the genotype produced from the assay (or analytic) accurate?
- Clinical validity – is the risk score accurate?
- Clinical utility – is the test useful in a clinical setting? Do individuals change their behavior?
- Physicians are not equipped
- Professional access
- Regulation
- Security/Privacy
- Long term effect on genetic research/Commercial exploitation

The Great Debate

Analytic validity – is the genotype produced from the assay (or analytic) accurate?

Navigenics' Approach

- Analytic accuracy of the overall Affy6.0 platform (1 million assays) is >99.5%
- Analytic accuracy of a particular SNP assay that we use is >99.9% as determined by genotype concordance in both DNA from cell lines (n=270) and paired blood/saliva samples (n=66)
- CLIA certified lab ensures quality laboratory testing and is regulated by CMS, lab required to undergo proficiency testing
- Additional measures ensure sample integrity:
  - Double barcode the saliva collection kit and return packaging ensure samples collected at the same address did not get mixed up
  - Member-specific barcoding
  - Sample identity double checked by confirmed platform determined gender with member-entered gender
  - Automated robotic processing to minimize sample mix-ups
  - Any sample run more than once undergoes additional QC steps to ensure identical results

The Great Debate

Clinical validity – is the risk score accurate?

- Navigenics starts with the average, gender-specific risk for the general population which incorporates both genetic and non-genetic risk factors. Then, we show members how that risk changes based on their known genetic risk factors that we test for. We clearly display how much of a condition is caused by genetic factors compared with non-genetic factors. We believe there is significant information that is helpful to our members especially when the interplay of environmental risk factors is also mentioned.
- Navigenics clearly states that we don't test for particular low-frequency variants that run in families and greatly increase the risk of disease. Customers are advised to discuss any conditions that occur in their family with their genetic counselor and may be referred for further testing, as appropriate.
- The federally funded human genome project was completed over 7 years ago, and yet the public has yet to see broadly applicable benefits of this research. While we realize that risk estimates will change over time, we believe that partial information is better than no information particularly when only protective health behaviors will result. Navigenics clearly states in the product that the risk score is based on currently available genetic evidence.

The Great Debate

Clinical validity – is the risk score accurate?

- Longitudinal studies follow a group of individuals over time and see if they develop the disease, and thus create a less artificial environment than case/control studies to measure the effect of genetic risk factors. However, there is evidence showing that the genetic risk factors identified in case/control studies show similar levels of risk when applied to longitudinal cohort study. *See Florez NEJM 355:241. 2006, Florez Diabetologia 51:451. 2008.*
- There is increasing evidence that amalgamated risk scores generated from genetic risk factors each with a small effect can impact the overall risk of disease. These studies have shown receiver/operator characteristics that are significantly better than random with AUC measurements between 60%-80%. In some cases these risk scores offer improvement over currently used risk assessment tools. *See Gold NatGen 38:458 2006, Weedon PlosMed 3:e374 2006, Morrison AJE 166:28 2007, Lu AJHG 82:641 2008, Kathiresan NEJM 358:12 2008*
- Navigenics only uses genetic risk factors that show consistent statistically significant effect sizes in multiple sample populations of the same ethnicity. Furthermore, each association is evaluated against a set of curation criteria that specifically address many of the pitfalls that have plagued the genetic association literature. *See Hirschhorn J Clin Endocrinol Metab 87:4438 2002, Hirschhorn Genet Med 4:45 2002, Cardon Nat Rev Genet 2:91 2001, Altshuler NEJM 338:1626 1998, Ioannidis Nat Genet 29: 306 2001, Dahlman Nat Genet 30: 149 2002, NCI-NHGRI Working Group on Replication in Association Studies. Nature 447:655 2007.*

The Great Debate

Clinical utility – is the test useful in a clinical setting? Do individuals change their behavior?

- This issue clearly illustrates the different approaches of medicine and public health. In medicine, a physician conducts tests and intervention that are in the best interest of his/her individual patients, while public health policies are intended to promote maximum benefit for the maximum number of individuals in a cost-effective manner. Since the Navigenics product is purchased by an individual, we feel that the cost/benefit discussion is not appropriate.
- We agree that there is a poor success record in developing and implementing effective disease prevention strategies. It is thought that one of the reasons is that the strategies suggested are not personalized (Syme Soc & Prev Med 51:247 2006). By providing personalized genetic and environmental risk factors we believe that individuals will change their behavior. Recent evidence from the REVEAL study on how individuals alter their behavior in response to genetic testing for Alzheimer's disease (Alzheimer Dis Assoc Disord 22:94 Jan/Mar 2008) has shown that individual do in fact alter their behavior based on their test results.
- Navigenics focuses on only actionable medical conditions that can be affected by lifestyle changes, early screening, increased awareness, and available medicine or treatment options. We always encourage customers to adopt healthy behaviors even if they show a decreased genetic load since a significant proportion of the disease is affected by environmental factors.
- Navigenics believes that we don't need to wait for definitive medical studies in order to show that behavior modification decreases disease incidence among genetically loaded individuals. The current medical evidence from randomized clinical trials that don't include genotype do show that behavior modification can reduce disease risk and we see no reason to believe that genotype confounds this association.
- Furthermore, Navigenics believes that individuals should be given the opportunity to change their behavior.
- It is true that some individuals may experience anxiety as a result of learning their genetic load, so we provide access to genetic counseling for all members. For individuals with very high risk we provide additional encouragement for them to contact their genetic counselor if the wish.

The Great Debate

Security / Privacy

- We operate in a HIPAA consistent manner
- We require opt-in for internal research and/or third party research
- Privacy and security policies ensure that our members can feel comfortable and confident receiving genetic information and analyses, and that they alone control how that information is to be used and distributed.
- We use the most advanced data protection systems; we safeguard, maintain and update your genetic profile in a highly secure environment. All customer profiles are anonymous to assure data security.
- Although there is concern about insurance companies misusing genetic information, there are currently no cases on record of this happening. We are very diligent about communicating how to avoid this problem to our members.

The Great Debate

Long-term effect on genetic research / Commercial exploitation

- Transparency in what we are testing for, assumptions in our risk score calculations, statements about the state of the science
- Informed consent is required
- We are taking a responsible approach – providing information about medically relevant conditions that are socially responsible (excluding HIV resistance, for example)
- We will not sell our member's genetic information in any way
- Individuals can opt-in to donate their genotype information to our product refinement efforts and our prospective outcomes trials research.

PMC Industry Standards Setting Conference

- Announced April 8th, in partnership with the Personalized Medicine Coalition (www.**personalizedmedicinecoalition**.org)
- Public conference event in Dec 2008, Washington D.C. venue TBD
- Broad participation of key stakeholders
- Potential Areas of focus for dialogue and recommendations:
  - Implementation of Privacy Protections for Online data
  - Operational/Lab Processing Standards
  - Diffusion of Communication Methods for Risk-based Information
  - Ensuring Consumers Understand and Adopt Genetic Risk-based information
  - Assessing Clinical Validity of Association Studies
  - Defining Actionable Health Information
  - Educating the Provider and Public

Long-term effect on genetic research / Commercial exploitation

- Transparency in what we are testing for, assumptions in our risk score calculations, statements about the state of the science
- Informed consent is required
- We are taking a responsible approach – providing information about medically relevant conditions that are socially responsible (excluding HIV resistance, for example)
- We will not sell our member's genetic information in any way
- Individuals can opt-in to donate their genotype information to our product refinement efforts and our prospective outcomes trials research.

Regulation

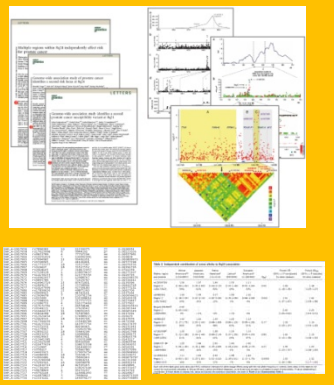
- We are in proactive discussions with relevant regulatory agencies to assist in developing appropriate regulatory standards for the industry.
- As our deep diligence tells us that we operate in a manner consistent with currently applicable regulatory guidelines.
- We support**ed** GINA!
- Informed consent is required and we do not test minors.
- We are completely transparent as to our scientific and clinical criteria, our calculations, and our primary references.
- We adhere to testing guidelines and position statements of professional organizations including the National Society of Genetic Counselors, the American College of Medical Genetics, and the American Society of Human Genetics.

Navigenics' competencies & partnerships

Core competencies

Platforms and assays

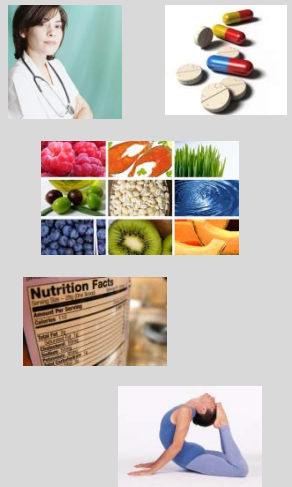
Scientific and Clinical Curation



Personalized Web Portal

Customer Experience

Lifestyle and Behavior



Partnerships

Correlagen

DNAdirect

DukeMedicine

AFFYMETRIX

Medscape from WebMD

Partnerships

MAYO CLINIC

PARTNERS HEALTHCARE

Scripps

I would like to communicate to you that …

- We are facing a health care crisis from CCND in this generation – and prevention is the only feasible solution
- Validated “genetic risk factors” are not so different than validated environmental risk factors, and can be used to practice the “art of medicine” in the identical fashion
- Genetic risk factors can be used to refine risk and drive additional focused prevention behaviors and early detection paradigms
- Delivery of the information in an accurate and private fashion to the public is necessary to meet timelines

Genetic and Epidemiology Team

David Botstein, MD, PhD
Michele Cargill, PhD
Eran Halperin, PhD
Shannon Kieren, MS, CGC
Isaac Kohane, MD, PhD
Elissa Levin, MS, CGC
Michael Nirenberg, MD
Badri Pakhukasahasram, PhD
Nik Schork, PhD
Elana Silver, MPH
Daryl Thomas, PhD
Heather Trumblower, MS
Jeffrey Trent, PhD
Vance Vanier, MD
Jennifer Wessel, PhD, MPH