From Linkage Maps to Quantitative Trait Loci: The History and Science of the Utah Genetic Reference Project

Stephen M. Prescott,[1] Jean Marc Lalouel,[2] and Mark Leppert[2]

[1]Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104; email: Steve-Prescott@omrf.ouhsc.edu

[2]Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112; email: jml@genetics.utah.edu; mleppert@genetics.utah.edu

Annu. Rev. Genomics Hum. Genet. 2008. 9:347–58

First published online as a Review in Advance on June 10, 2008

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

This article's doi: 10.1146/annurev.genom.9.081307.164441

1527-8204/08/0922-0347$20.00

Key Words

genetics, polymorphism, mapping, CEPH

Abstract

One of the early decisions in what became the Human Genome Project was to recruit families that would serve as a reference set, thereby focusing efforts to create human genetic maps on the same sets of DNA samples. The families recruited from Utah provided the most widely used samples in the Centre d'Etudes du Polymorphisme Humain (CEPH) set, were instrumental in generating human linkage maps, and often serve as the benchmark for establishing allele frequency when a new variant is identified. In addition, the immortalized cell lines created from the peripheral blood cells of these subjects are a broadly used resource and have yielded insights in many areas, from the genetics of gene expression to the regulation of telomeres. More recently, these families were recontacted and underwent extensive, protocol-based evaluation to create a phenotypic database, which will aid in the study of the genetic basis of quantitative traits. As with the earlier efforts, this project involved collaborations among many investigators and has yielded insights into multiple traits.

ORIGINS OF THE HUMAN LINKAGE MAP

Until the late 1970s, a limited set of protein polymorphisms (~30) was available to perform linkage studies in families in which a single mendelian trait, generally an inherited disorder of metabolism, was segregating (25). Typically, linkage studies relied on the tests of cosegregation between a trait behaving as a single mendelian and a genetic marker; these tests had been developed and first applied by Newton Morton (23, 24). The likelihood of success was low, given the limited coverage of the genome afforded by the markers then available.

New technologies emerging from molecular biology that afford the definition and use of an unlimited number of genetic markers, together with the development of algorithms to model segregation of multiple loci, have revolutionized the field of genetic mapping. A University of Utah–sponsored genetics retreat at Alta Summit, Utah, in April 1978 changed the view of how to approach the creation of a human linkage map and laid the foundation to map systematically the human genome. Discussions at this meeting culminated in the paper by David Botstein, Raymond White, Mark Skolnick, and Ronald Davis (3) in which they proposed the use of randomly derived restriction fragment polymorphisms (RFLPs) to generate a linkage map of the human genome. Because RFLPs had already been identified in human mitochondrial DNA and DNA from several human cell lines (14, 27), the authors predicted that prior knowledge of specific gene sequences was not necessary to generate this type of linkage map. Botstein and colleagues also suggested that only 150 RFLPs were needed to generate the linkage map; however, these RFLPs should be of high quality and be highly informative. To establish linkage at any given locus, Botstein and colleagues calculated that DNA from several hundred individuals was needed and concurred with Thompson and coworkers (28) that DNA collected from multigenerational families was more useful than DNA from nuclear families. Thus, they proposed a blind search for RFLPs in DNA from healthy individuals in extended Utah pedigrees (3).

The discovery of RFLPs and their use in linkage mapping efforts proceeded at a rapid pace; early successes with well-defined mendelian disorders amply demonstrated the power of the linkage strategy using these markers. Indeed, James Gusella from Harvard Medical School, working with a team from Venezuela (12), found an RFLP on chromosome 4 that linked to the Huntington's disease gene. Additional studies yielded DNA markers in genetic linkage with Duchenne (26) and Becker (16) muscular dystrophies, X-linked retinitis pigmentosa (2), and fragile X–linked mental retardation (5).

IMPLEMENTATION OF THE CEPH COLLABORATION

As multiple genetic research groups became quite active in this area, it soon became apparent that the efficiency of genetic mapping using this wealth of new RFLP markers could be markedly enhanced if these markers were characterized on a common source of DNA. Indeed, genetic linkage relationships between RFLPs developed and genotyped in distinct pedigrees at various laboratories could not readily be assembled into linear genetic maps of the human chromosomes using the pooled genotypic information.

Jean Dausset, who had won the 1980 Nobel Prize in Physiology or Medicine for his work on human leukocyte antigen (HLA) and received a large private donation, proposed in 1983 to develop the Center d'Etude du Polymorphisme Humain (CEPH) (11) as a nonprofit research institute to fill this specific gap. He had conducted his work on histocompatibility antigens and their roles in graft rejection through the participation of dedicated families of volunteers with large sibship size. He understood the value that a common typing panel would have in genetic mapping and wished to contribute his resources to the international scientific community. In concert with Daniel Cohen and Jean-Marc Lalouel, a program was developed

that began with formal invitations from CEPH to genetic researchers from around the world to participate in the first major collaborative research project in human genetics, the first international genome project. CEPH proposed to act as the centralized clearing house to produce and supply DNA from a common panel of reference families, collate genotyping data from participating laboratories into a central database, and redistribute these data to collaborators.

The key premise that a "human genetic map will be efficiently achieved by collaborative research on DNA from the same sample of families" (8) was enthusiastically endorsed by leading geneticists engaged in gene mapping at the time. Of particular significance to CEPH was the participation of Ray White, one of the authors of the key concept paper that introduced RFLPs to gene mapping (3), and his research team in Salt Lake City. Ray White had earnestly joined the project with extended pedigrees and kindreds with large sibship size collected in Salt Lake City and was leading a major effort to identify new RFLPs. His participation was crucial for the success of the CEPH initiative. Daniel Cohen and Jean-Marc Lalouel met with Ray White at a Miami Symposium in early 1984. Ray White generously agreed to participate and contribute his family resources as necessary; leading concepts of the collaboration were delineated in this productive encounter. They agreed that specifics of the CEPH operational features would be defined in the course of a meeting of collaborators from the international community sponsored by CEPH in Paris later in the year.

The meeting occurred in Paris on October 20, 1984. A sample of planning notes from Lalouel's notebook is shown in **Figure 1**. Howard Cann had left Stanford University to join CEPH permanently. Scientific, strategic, and logistic issues were discussed and settled in the course of this meeting that set the stage for the future of a most successful international endeavor to generate a linkage map of the human chromosomes. A set of 40 families was selected as a reference panel on the basis of the consensus that emerged among collaborators. Emphasis was put on nuclear families with large sibships (mean sibship size was 8.3) and their parents, and wherever feasible grandparents would be included because they would provide phase information. The large sibships afforded replication of segregating events within the family and the genetic knowledge of maternal and paternal grandparents provided the ability to determine phase of loci (19, 30). The families consisted of 10 French families contributed by Jean Dausset, 27 families contributed by Ray White, one Amish family, and the two core families from a large Huntington pedigree from Venezuela. The panel, evidently a compromise to achieve overall consensus, was subsequently extended by inclusion of an additional set of 20 Utah families with grandparents, so the latter were available in 44 of these 60 families. This extension was largely justified by the need to refine mapping with enhanced power once linkage with a particular mendelian disorder in select pedigrees was established. The higher the map density, the smaller the interval into which a candidate locus could be narrowed, thereby markedly enhancing the power of positional cloning.

Other elements of the CEPH collaboration were specified at that time, including production and distribution issues for DNA, typing strategies, procedures for data collection and redistribution among collaborators, and protocols for publication. Much of this was briefly described years later in a review article (8).

Following this initial meeting, all reference lymphoblastoid cell lines from members of the panel were assembled and placed in culture at CEPH, DNA was produced in large quantities, and aliquots were forwarded to participating laboratories. In parallel, software was provided to share data through the CEPH database and the LINKAGE computer program (17, 18) was provided to all members to analyze genotypic data upon assembly into a defined format. This major undertaking, which required the establishment of large production facilities at CEPH to store cell lines, expand them in culture, and perform large-scale extraction of DNA,

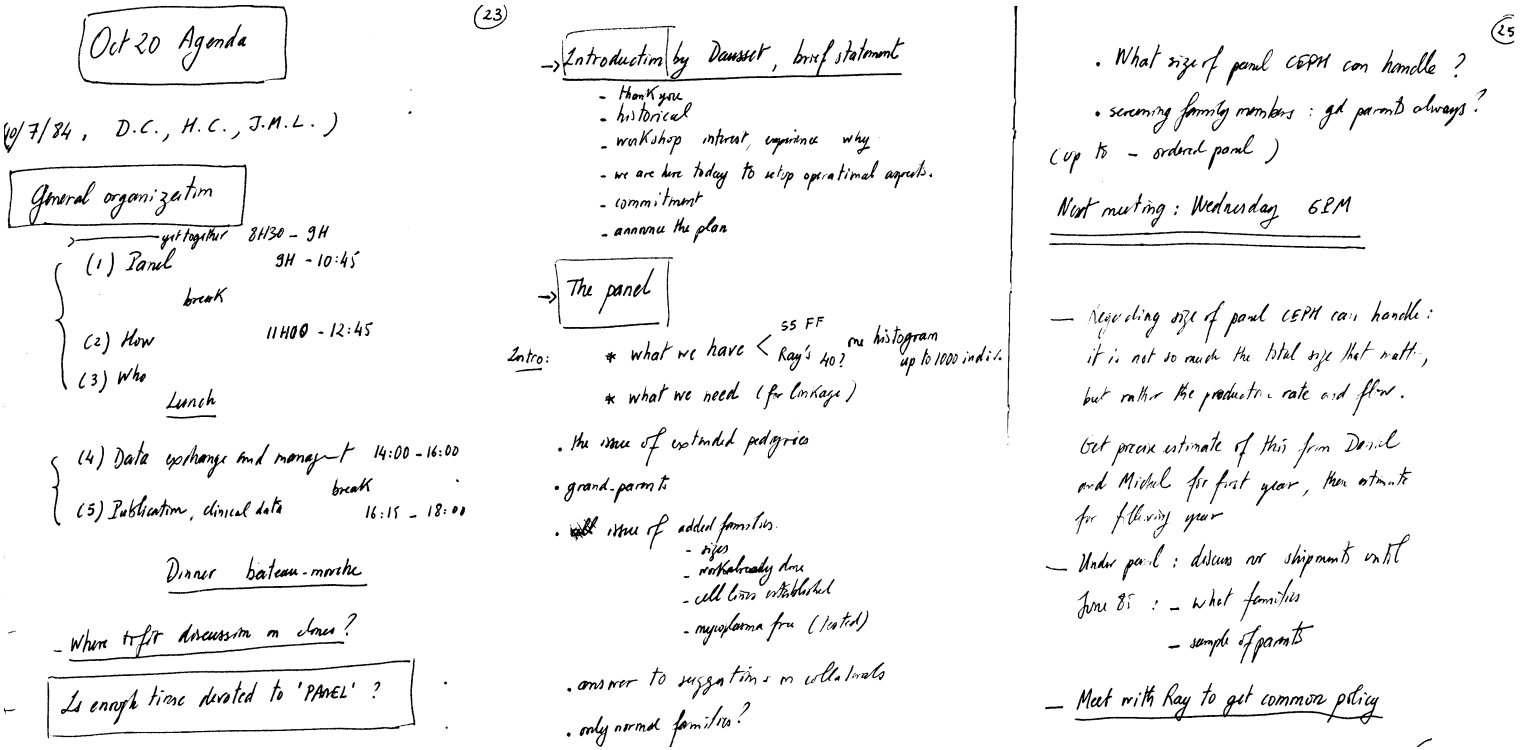

Figure 1

Original notes from Jean Marc Lalouel's notebook on the planning of the first Center d'Etude du Polymorphisme Humain (CEPH) Meeting.

proceeded smoothly, and the CEPH community expanded at a rapid pace to include 63 scientific laboratories by 1989.

LINKAGE MAPS OF THE HUMAN CHROMOSOMES

By 1987, with the support of the Howard Hughes Medical Institute, a complete set of genetic maps of the human chromosomes was assembled by Ray White and Jean-Marc Lalouel and their associates in Salt Lake City; these maps were made available to the scientific community in an internally produced mimeograph document handed out prior to publication to scientists participating to the Human Gene Mapping Conference held in Paris in 1987 (31). These maps were subsequently released in formal publications. Another set of genetic maps generated internally at Collaborative Research, Inc. was published independently the same year (9). Many investigators developed and published the results of their own efforts, which were often focused on particular chromosomes where linkage with specific mendelian disorders was established. One of the first comprehensive human genetic maps was published in 1998 (4). This map utilized 8000 microsatellite markers developed by three centers and constructed a map from eight CEPH families. This map further documented the difference in female versus male recombination across chromosomes, and the increased female recombination yielded an overall length (cM) ratio of female to male of 1.6 for autosomes. In 2003 a 3.9-cM resolution map was published using 56 of the CEPH families and nearly 3000 single nucleotide polymorphism (SNP) markers (20). This map compared favorably with other previously published linkage maps and the human physical map. The 3.9-cM resolution map showed broad trends in sex-specific recombination rates across all chromosomes and demonstrated that every autosome showed an increased recombination in females compared with males, whereas the average female-to-male length ratio (1.7:1 cM) varied by chromosome and specific chromosome region.

ACHIEVEMENTS OF CEPH

In retrospect it is clear that the CEPH project, which arose from the independent initiatives of investigators willing to advance a common goal, was the first successful international genome collaboration. This project markedly accelerated the production of genetic maps, which enabled investigators to map and thereafter proceed toward positional cloning of the unknown genes underlying a host of mendelian disorders. The resource was further enhanced by providing reference lymphoblastoid cell lines from a subset of the CEPH panel to the global scientific community through the Coriell Cell Repository in Camden, New Jersey. Currently, cell lines from 61 CEPH families are available to researchers from the repository. Forty-eight of these are from Utah CEPH families and include parents and for the most part all four members of the grandparental generation. A total of 665 cell lines from the Utah CEPH families are deposited in Coriell, with an average of 8.3 sibs per family. Details for all CEPH families with pedigrees can be found on the Coriell Institute for Medical Research Web site at **http://www.coriell.org.**

The CEPH genotype database V10.0 (November 2004) now contains genotypes for 32,356 genetic markers, 21,480 bi-allelic markers, and 9900 microsatellite markers. This resource has more than 6 million genotypes and is available at **http://www.cephb.fr.**

The CEPH collaboration can claim several long-lasting cultural achievements in science, including a data-sharing strategy among the CEPH collaborators that used a single format and is still in use today. In addition, the ethical standards developed for the treatment of the confidential information for the CEPH families remain intact after 25 years; these standards were such a success that when asked to participate in the next phase of data collection most families readily agreed.

The CEPH panel also facilitated at least two subsequent collaborative endeavors. One consisted of using this genotypic resource to investigate extensive phenotypes of members of the panel as part of the Utah Genetic

Reference Project, described below. The other was to provide the HapMap project with reference sets of offspring and parents trios of Caucasian ancestry (7). The reconsent process of individuals for use of these cell lines in the HapMap project was facilitated by the continual contact of the Utah investigators with the original Utah CEPH families through the Utah Genetic Reference Project.

THE UTAH GENETIC REFERENCE PROJECT

By the early 1990s, genotypic data were known for more than 300 individuals from the 47 original CEPH families from Utah, and linkage maps had been generated for most of the human genome (11). But more value could be extracted from these data through the development of phenotypic information from the panel members that originated from Utah and southern Idaho.

Examination of the original institutional review board (IRB)-approved informed consent documents for these 47 families revealed that family members had provided informed consent to be recontacted for further studies. We decided to capitalize on the plethora of genetic information already obtained from these families and began a process to collect phenotypic data; thus began the Keck/Utah Genetic Reference Project (UGRP). At the time, we reasoned that this type of study at worst would result in a clinical analysis with no molecular basis and at best would provide a powerful tool to begin to understand the genetics of quantitative traits.

Because the original goals of the UGRP project were to identify the genes involved in normal variation in humans and to ascertain the genes involved in the predisposition to common human diseases, this was a high-risk proposal that could not readily be supported through conventional public agencies. In 1995, we sought funding from the Keck Foundation to begin to recontact 16 of the original 47 Utah families.

Initial funding by the Keck Foundation for the first phase of the phenotypic characterization assisted in the development of the infrastructure, the identification of the phenotypic characteristics to be measured, and the determination of the feasibility for such an undertaking. More than 50 local, national, and international scientists and clinicians were recruited during the planning stages for the development of testing and information planning protocols. These collaborators were recruited because they had expertise in ethics and consent issues or in clinically and scientifically relevant fields, including autoimmune, eye, pulmonary, or cardiovascular diseases; metabolic disorders; hematology; speech pathology; neuropsychiatric disorders; or cancer. These collaborators helped to develop an 85-page questionnaire that each family answered, administered and monitored the physical testing of the family members, conducted testing on biological samples, interpreted test data, and analyzed genetic data for linkages with quantitative traits and intermediate phenotypes.

Of the multitude of biochemical, clinical, and physical characteristics that could be measured, a final 180 characteristics were chosen on the basis of ease of data collection, cost of performing the test, clinical significance, and overall scientific interest (**Table 1**). These CEPH families were originally considered 'normal' because they were selected on the basis of the lack of phenotypic symptoms for genetic diseases. However, we did expect that rates of common diseases and traits would be normal in this population.

We initiated the recontact of the original families in 1997. By the time the original families were recontacted and asked to participate, many of the original grandparents had passed away, but a fourth generation had begun. Most families were both eager to participate and to include the fourth generation in our studies.

UTAH GENETIC REFERENCE PROJECT FAMILY REUNIONS

The most efficient method to rigorously collect the data in which we were interested in was to host a reunion of sorts for each participating

Table 1 Phenotypes collected from Utah Genetic Reference Project (UGRP)/Utah Center d'Etude du Polymorphisme Humain (CEPH) families

Vital signs at rest
Oxygen saturation by pulse oximetry
Blood pressure/urine and blood hormones and metabolites
Catecholamine metabolism
Pulmonary function tests [forced expiratory volume (FEV) in one second and forced vital capacity (FVC)]
Attention span, attention deficit disorder [test of variables of attention (TOVA) on the computer]
Bone density, indirect muscle mass measurements
Taste and smell
Handedness
Voice pitch, vocal senescence amplitude, and frequency measurements
Audiology testing
Facial photography and facial/cranial/cranial 3-D laser scan
Dental exam with teeth impressions
Electrocardiogram
Echocardiography
Ophthalmological testing, including accommodation and visual acuity
Reflectometric assay of skin melatonin content and composition
Lipids, lipoproteins by subtype
Telomere length and aging
Susceptibility to various common human viruses
Blood coagulation proteins
White cell subtypes, innate immunity parameters
Personality traits
Anthropometric measurements

family. Each family reunion was planned for a weekend. On Friday evening, the family arrived at a Salt Lake City hotel, received dinner, and met with the study coordinator and Drs. Mark Leppert and Andy Peiffer to ask any questions and provide informed consent. On Saturday, the family was transported to the University of Utah General Clinical Research Center and submitted to a day-long collection of quantitative and qualitative data (**Table 1**). We arranged the testing schedule such that fasting blood samples were collected from each family member before 9:30 AM, after which the participants were fed breakfast and escorted to the various locations within the General Clinical Research Center to finish their testing. Study collaborators or representatives also attended these clinic days to oversee the collection of the data specific to their respective projects that were outside the scope of our funding. All data collection concluded by 5:00 PM, and family members who were not local residents were provided with another night's stay at the hotel. Collection of all measurements and samples was carried out at the same time and followed an identical protocol for all family visits to ensure consistency in phenotypic measurement between families.

By 1999, we secured a second round of funding from the Keck Foundation to support recontact and data collection for the remaining families that originally participated in the CEPH database. Amazingly, 42 out of the original 47 CEPH families agreed to participate in the UGRP; one family refused to participate and four families were lost to follow-up. Data from the last family were collected in September of 2005. At this time new biosamples collected from fasting family members included plasma and DNA from peripheral blood; no new lymphoblastoid cell lines were established. A set of 537 microsatellite markers was genotyped on these 42 families with high heterozygosity, with an average of 75%, by laboratories at Utah and Marshfield, WI. Very recently, all

Table 2 Heritabilities of selected anthropomorphic measures in 42 Utah Genetic Reference Project families

Variable name	Heritability
Body mass index	0.30
Height	0.97
Elbow breadth	0.72
Biacromal diameter	0.76
Biygomatic diameter	0.70
Face height	0.71
Height anterior-superior iliac spine	0.90
Sitting height/crown-rump	0.84
Stature/supine length	0.96
Total arm length	0.79

members of the three-generation UGRP were genotyped on the Affymetrix SNP array 6.0 chip.

HERITABILITY ESTIMATES FROM 42 UTAH GENETIC REFERENCE PROJECT FAMILIES

Prior to any genetic linkage analysis, we estimated heritabilities of quantitative traits to assess the role of genes for a given phenotype. As an example, **Table 2** provides a list of select anthropomophric measurements that document high levels of heritabilities in the 42 UGRP families. These heritability estimates were computed using the polygenic function available in the SOLAR program (1). All measures were adjusted for gender, age, age^2, and age^3. In the following sections, a few practical applications are provided to document the significance and practical usefulness of the UGRP program.

MAPPING QUANTITATIVE TRAIT LOCI FOR LYMPHOCYTE SUBPOPULATIONS

Levels of defined lymphocyte subpopulations are commonly used in the prognosis and monitoring of a variety of human diseases. These subpopulations can be separated readily by fluorescence-activated cell sorting (FACS) using epitopes that tag each subset of cells, affording counts of total lymphocytes, CD4 T cells, CD8 T cells, the CD4/CD8 cell ratio, CD19 B cells, and natural killer cells (21). This investigation was performed in each member of the first 15 CEPH families included in the UGRP project to ascertain the genetic variation among these phenotypes. The calculated heritabilities ranged from 0.46 to 0.61, consistent with a substantial genetic component. To test for major effects, a whole-genome scan was performed that identified significant quantitative trait loci (QTL) on chromosomes 1, 2, 3, 4, 8, 9, 11, 12, and 18. Each QTL accounted for a significant proportion of the phenotypic variance of lymphocyte subpopulations (13). To test whether candidate genes underlying such QTL could be identified, the chromosome 18 QTL for CD4 T cells was selected because it encompasses genes implicated in T cell function, particularly *Bcl-2*. A multiallelic short tandem repeat polymorphism (STR) identified at this locus was used in further tests of association, and high significance was obtained for one allele in this series. This work provided clear proof of principle that QTL mapping could be applied in these families as a first step toward the identification of genes underlying common genetic variation.

PHENYLTHIOCARBAMIDE TASTING

The inheritance of the ability to taste phenylthiocarbamide (PTC) was long considered to be a classical problem in human genetics because it did not quite follow Mendelian rules for a single locus. In the late 1980s some groups suggested that a multifactorial effect could explain the trait. Previous studies analyzed linkage data based on taster versus nontaster status using a demonstrated threshold. These data indicated strong support for linkage to the Kell blood group antigen and other chromosomal regions, working under the assumption of recessive inheritance. UGRP families were asked to taste a range of 14 PTC concentrations starting at 1 μM and increasing twofold to 8.54 mM, the results of which were ultimately presented

as a range of PTC scores from 0 (least sensitive to taste) to 14 (most sensitive). Analysis of the UGRP family members' ability to taste a range of concentrations of PTC indicated a surprising bimodal distribution of PTC scores (means 3.16 and 9.26; $\chi^2 = 93.27$, df $= 3$, $P < 0.001$). There was considerable variation in this phenotype, and the use of a threshold approach to analyze the data, regardless of where the threshold was placed, would inevitably misclassify some individuals. Linkage analyses using taste sensitivity as a quantitative trait revealed a major locus on chromosome 7q and a secondary locus on chromosome 16p (10). Subsequent to the localization of the major locus and additional fine mapping on chromosome 7q, a gene for PTC tasting was discovered in the families contributing to the linkage signal on chromosome 7q (15). This study identified a small region of strong linkage disequilbrium on chromosome 7q that contained a single gene, the TAS2R taste receptor gene. Further analysis led to the discovery of three missense SNPs that gave rise to five major haplotypes specifically associated with PTC taster or nontaster phenotypes. This study demonstrated the power of these UGRP families in the detection of quantitative trait loci with highly significant logarithm of the odds (LOD) scores even when locus heterogeneity was present. Moreover, the large number of UGRP family members included in the phenotypic test proved to be sufficient for purposes of gene discovery.

GENETICS OF GENE EXPRESSION

Warren and colleagues (29) used vaccine virus vectors for transient induction of HA-8, a minor histocompatibility antigen, in CEPH lymphoblastoid cell lines, followed by detection of the antigen with a cytotoxic T cell clone by standard in vitro cytotoxicity assays. By performing linkage analysis of surface markers on the Epstein-Barr virus (EBV)-transformed cell lines induced by this approach, these authors were able to map genes that control the induction of this histocompatibility antigen.

In a seminal experiment that would lay the stage for subsequent studies, a group led by Vogelstein and Kinzler (32) showed that allelic variation in human gene expression could be unraveled using CEPH reference cell lines as a source. After validation of their approach in a preliminary experiment, they proceeded to screen SNPs for 13 genes using RNA from 96 unrelated CEPH founders. They applied a quantitative assay based on fluorescent dideoxy terminators to quantitate the levels of each of two alleles at each locus. They found that 17 of 37 individuals were heterozygous for any given gene. Significant differences in allelic expression were observed for 6 of the 13 genes studied. They next examined the families of nine individuals who exhibited allelic variation. Three families were informative and displayed expression patterns consistent with Mendelian inheritance. The researchers concluded that *cis*-acting inherited variation in gene expression is relatively common among normal individuals.

Taking advantage of the high throughput afforded by microarrays to investigate gene expression profiles, two groups have independently used part of the CEPH panel to further establish that common genetic variation underlying gene expression could be mapped by linkage analysis. Cheung and Spielman and their associates (22) investigated mRNA from lymphoblastoid cell lines of 14 Utah CEPH families obtained from Coriell. Using Affymetrix Genome Focus Arrays, they characterized gene expression for 8500 genes in 94 unrelated CEPH grandparents and restricted subsequent analysis to 3554 genes that exhibited greater variation between individuals than between replicates. Expression profiling and genotypes at 2756 SNP loci previously mapped in 56 CEPH reference families (20) afforded linkage mapping of 142 to 984 expression profiles, depending on the stringency of the linkage detection test. Linkage occurred with markers in the genomic region both encompassing or not encompassing the gene coding for the transcript examined, suggesting *cis* or *trans* effects. In further tests of 17 genes that exhibited linkage in *cis* under the most stringent conditions,

SNP typing within or near the target gene revealed significant association by quantitative transmission disequilibrium test (QTDT) in 14 (82%) of these genes, strongly supporting the hypothesis that a common variant in the region accounts for the differential gene expression.

Similar work was pursued by Eric Schadt and colleagues (21) in 15 Utah CEPH families also obtained from Coriell. Expression profiling was performed for 23,499 genes, of which 2430 passed a criterion for differential expression in at least half of the offspring tested. Linkage analysis was performed using 346 autosomal markers obtained from the CEPH database (version 9.0) and yielded significant linkage for 333 to 132 phenotypes, depending on the stringency of the pointwise significance selected. In this study, 13 of the 333 and 25 of the 132 mapped expression profiles could be classified as occurring in *cis*. The differences between the two studies were essentially methodological in nature. Taken together, these data provided strong evidence that common genetic variation underlying gene expression can be detected by linkage analysis in CEPH reference families.

AGING AND TELOMERE LENGTH

A recent study (6) demonstrated a significant association ($p = 0.004$) between telomere length in white blood cells and mortality in people aged 60 years or older. This study utilized biosamples from the original Utah CEPH families and cause of death information from the Utah Population Database and the Social Security death index. Individuals with shorter telomeres had a mortality rate of nearly twice that of those with longer telomeres, and individuals from the bottom half of the telomere length distribution had a heart disease mortality rate that was more than three times that of those in the top half of the distribution. The benefit of continued connection to the families over many decades for aging studies is obvious. Another potential advantage of conducting a long-term longitudinal study with a large cohort of families is the ability to collect DNA samples from individuals over the course of several decades and then compare telomere length decline with mortality.

SUMMARY AND FUTURE DIRECTIONS

The efforts of the Utah Genetics Reference Project have affected multiple areas of genetics. The initial goal, which was to create a resource for the generation of linkage maps early in the human genome project, was successful from the outset. The availability of DNA samples from three-generation families with large sibships and grandparents fulfilled needs throughout the different stages of map making. It was a pleasant surprise that the effort to collect these samples also resulted in additional immediate benefits. Indeed, these individuals, via the DNA samples and the immortalized cell lines they generously contributed, became a worldwide reference resource as investigators seeking to find mutant alleles associated with disease used this set of DNA to determine the frequencies of alleles in a normal population. Subsequently, it was obvious that this was a key resource to be used in the HapMap project.

The initial decision to create immortalized cell lines proved to be prescient as these lines have had multiple uses beyond the initial goal of serving as a permanent source of DNA. The cell lines have been used for studies of the genetic basis for gene expression, tissue transplantation, and genomic stability over time in cell culture, among other efforts.

In the most recent phase of this project, the rerecruitment of the subjects for phenotypic analysis to study the genetic basis of quantitative traits has likewise yielded results that highlight the value of the resource. The approach, which was to make precise measurements according to a rigorously standardized protocol, offered advantages over studies that have used clinical records because the data in those circumstances may be flawed as a result of variation in measurement methods. Additionally, the decision to treat all traits as continuous variables added an important dimension to the power of the project.

One crucial element for the success of this project that has extended for more than 25 years has been the dedicated involvement of the families. It is interesting to note that they are tremendously proud of their contributions. They are informed about the confidential use of their data through regular communications (including a newsletter), and although few if any of them have scientific training, they have followed the studies with interest and pride. Likewise it is a source of pride to the investigators that the confidentiality and privacy of the subjects have been maintained over this long period of time and with many collaborators. This provides reassurance that large scale, longitudinal projects with genetic information can be conducted in a manner that does not harm the participants.

In conclusion, the Utah Reference Genetics Project offers a distinct example of the utility of a long-range, multifaceted study of large multigenerational families to unravel genetic variation underlying health and disease.

DISCLOSURE STATEMENT

The authors are not aware of any biases that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

The investigation described herein was supported by a Public Health Services research grant to the Huntsman General Clinical Research Center at the University of Utah, grant number M01-RR00064, from the National Center for Research Resources. Research was also supported by generous gifts from the W.M. Keck Foundation and from the George S. and Delores Doré Eccles Foundation. We would like to extend our sincere thanks to all family members who participated in the Utah Genetic Reference Project. Thanks also to Andreas P. Peiffer, M.D., Ph.D., UGRP Medical Director, and Melissa M. Dixon, UGRP Study Coordinator.

LITERATURE CITED

1. Almasy L, Blangero J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* 62:1198–211
2. Bhattacharya SS, Wright AF, Clayton JF, Price WH, Phillips CI, et al. 1984. Close genetic linkage between X-linked retinitis pigmentosa and a restriction fragment length polymorphism identified by recombinant DNA probe L1.28. *Nature* 309:253–55
3. Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32:314–31
4. Broman KW, Murray JC, Sheffield VC, White RL, Weber JL. 1998. Comprehensive human genetic maps: Individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* 63:861–69
5. Camerino G, Mattei MG, Mattei JF, Jaye M, Mandel JL. 1983. Close linkage of fragile X-mental retardation syndrome to haemophilia B and transmission through a normal male. *Nature* 306:701–4
6. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA. 2003. Association between telomere length in blood and mortality in people aged 60 years or older. *Lancet* 361:393–95
7. Consort. The Int. HapMap. 2005. A haplotype map of the human genome. *Nature* 437:1299–320
8. Dausset J, Cann H, Cohen D, Lathrop M, Lalouel JM, White R. 1990. Centre d'etude du polymorphisme humain (CEPH): collaborative genetic mapping of the human genome. *Genomics* 6:575–77
9. Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, et al. 1987. A genetic linkage map of the human genome. *Cell* 51:319–37
10. Drayna D, Coon H, Kim UK, Elsner T, Cromer K, et al. 2003. Genetic analysis of a complex trait in the Utah Genetic Reference Project: a major locus for PTC taste ability on chromosome 7q and a secondary locus on chromosome 16p. *Hum. Genet.* 112:567–72

11. Group, NIH/CEPH Collab. Mapp. 1992. A comprehensive genetic linkage map of the human genome. *Science* 258:67–86
12. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, et al. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306:234–38
13. Hall MA, Norman PJ, Thiel B, Tiwari H, Peiffer A, et al. 2002. Quantitative trait loci on chromosomes 1, 2, 3, 4, 8, 9, 11, 12, and 18 control variation in levels of T and B lymphocyte subpopulations. *Am. J. Hum. Genet.* 70:1172–82
14. Hutchison CA 3rd, Newbold JE, Potter SS, Edgell MH. 1974. Maternal inheritance of mammalian mitochondrial DNA. *Nature* 251:536–38
15. Kim UK, Jorgenson E, Coon H, Leppert M, Risch N, Drayna D. 2003. Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science* 299:1221–25
16. Kingston HM, Thomas NS, Pearson PL, Sarfarazi M, Harper PS. 1983. Genetic linkage between Becker muscular dystrophy and a polymorphic DNA sequence on the short arm of the X chromosome. *J. Med. Genet.* 20:255–58
17. Lathrop GM, Lalouel JM. 1984. Easy calculations of LOD scores and genetic risks on small computers. *Am. J. Hum. Genet.* 36:460–65
18. Lathrop GM, Lalouel JM, Julier C, Ott J. 1984. Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci. USA* 81:3443–46
19. Malhotra A, Cromer K, Leppert MF, Hasstedt SJ. 2005. The power to detect genetic linkage for quantitative traits in the Utah CEPH pedigrees. *J. Hum. Genet.* 50:69–75
20. Matise TC, Sachidanandam R, Clark AG, Kruglyak L, Wijsman E, et al. 2003. A 3.9-centimorgan-resolution human single-nucleotide polymorphism linkage map and screening set. *Am. J. Hum. Genet.* 73:271–84
21. Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, et al. 2004. Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* 75:1094–105
22. Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, et al. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature* 430:743–47
23. Morton NE. 1955. Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* 7:277–318
24. Morton NE. 1956. The detection and estimation of linkage between the genes for elliptocytosis and the Rh blood type. *Am. J. Hum. Genet.* 8:80–96
25. Mourant AE, Kopeć AC, Domaniewska-Sobczak K. 1976. *The Distribution of the Human Blood Groups and Other Polymorphisms*. London: Oxford Univ. Press
26. Murray JM, Davies KE, Harper PS, Meredith L, Mueller CR, Williamson R. 1982. Linkage relationship of a cloned DNA sequence on the short arm of the X chromosome to Duchenne muscular dystrophy. *Nature* 300:69–71
27. Potter SS, Newbold JE, Hutchison CA 3rd, Edgell MH. 1975. Specific cleavage analysis of mammalian mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 72:4496–500
28. Thompson EA. 1978. Linkage and the power of a pedigree structure. In *Genetic Epidemiology*, ed. NE Morton, CS Chung, pp. 247–53. New York: Academic
29. Warren EH, Otterud BE, Linterman RW, Brickner AG, Engelhard VH, et al. 2002. Feasibility of using genetic linkage analysis to identify the genes encoding T cell-defined minor histocompatibility antigens. *Tissue Antigens* 59:293–303
30. White R, Leppert M, O'Connell P, Nakamura Y, Julier C, et al. 1986. Construction of human genetic linkage maps: I. Progress and perspectives. *Cold Spring Harbor Symp. Quant. Biol.* 51(Pt. 1):29–38
31. White R, Lalouel JM, Leppert M, O'Connell P, Nakamura Y, Lathrop GM. 1989. Linkage maps of human chromosomes. *Genome* 31:1066–72
32. Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. 2002. Allelic variation in human gene expression. *Science* 297:1143

Annual Review of Genomics and Human Genetics

Volume 9, 2008

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Genomics and Human Genetics* articles may be found at http://genom.annualreviews.org/