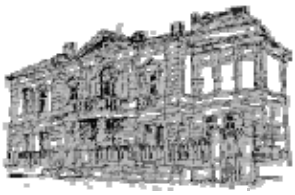
THE NORWEGIAN ACADEMY OF SCIENCE AND LETTERS



A B S T R A C T S

"Genomics and Genetic Applications in Medicine"

Oslo, May 19 - May 21, 2005

organized by the Sig. K. Thoresen Foundation and
The Norwegian Academy of Sciences

Comparative Genomics at the Extremes

Eddy Rubin,
DOE Joint Genome Institute, USA

Two areas exploiting comparative genomic data will be discussed: the analysis in transgenic mice of a complete chromosomal set of enhancers identified through human fish sequence comparisons and the genomic sequencing and analysis of DNA isolated from 40,000 year old cave bear bones. A significant challenge exists in identifying and characterizing the complete set of gene enhancers present in the human genome. To perform large-scale assessment of this activity we have characterized all ~100 highly conserved human-fugu noncoding elements on human chromosome 16 using a high throughput mouse transgenic reporter assay.

In second study, we describe the construction, deep sequencing and analysis of two genomic libraries by direct cloning of unamplified DNA extracted from bones of two different 40,000-year-old cave bears. These results demonstrate that analytically significant quantities of ancient genomic sequence can be obtained with a moderate amount of sequencing, and confirm the feasibility of ancient DNA genome sequencing programs.

Nobrega, M.A., Ovcharenko, I., Afzal, V., Rubin, E.M. Scanning Human Gene Deserts for Long-Range Enhancers. ***Science*** 302:413 (2003)

Boffelli D, Nobrega MA, Rubin EM. Comparative genomics at the vertebrate extremes. ***Nature Review Genetics***, (6):456-65 (2004).

Nobrega MA, Zhu Y, Plajzer-Frick I, Afzal V, Rubin EM. Megabase deletions of gene deserts result in viable mice.***Nature***. 2004 Oct 21;431(7011):988-93.

Analysis of Gene Copy Number Variation in Human Disease

Jonathan Sebat[1], Christa Martin[2], Vaidehi Jobanputra[5], Amanda Yonan[6], B. Lakshmi[1], Jennifer Troge[1], Joan Alexander[1], Maoyen Chi[1], Nicholas Navin[1], Robert Lucito[1], John Healy[1], James Hicks[1], Andrew Reiner[1], Kenny Ye[3], Dan Geschwind[4], Dorothy Warburton[5], David Ledbetter[2], T. Conrad Gilliam[6], Michael Wigler[1]

[1] *Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 1172*

[2] *Department Department of Human Genetics, Emory University School of Medicine, 615 Michael St., Suite 315, Atlanta, GA 30322*

[3] *Dept. of Applied Mathematics & Statistics, Stony Brook University, Stony Brook, NY 11794*

[4] *Department of Neurology, University of California Los Angeles, 710 Westwood Plaza, Los Angeles, CA 90095*

[5] *Columbia University Medical Center, Babies Hospital South, 3959 Broadway, New York, NY, 10032*

[6] *Columbia Genome Center, Columbia University, New York, NY 10032*

Based on work by our group and others it has been shown that large-scale copy number polymorphisms (CNPs) are a significant source of human genetic variation. In contrast to DNA sequence variants such as SNPs and microsattelite repeats, CNPs have not been well characterized. Much remains to be learned about the genomic locations, frequency and stability of these structural variants and their importance in human genetic disease. We have expanded our studies of normal genetic variation to include data on 200 individuals. We will discuss the findings of this study and discuss the implications of CNPs for genetic studies of disease.

The abundance of gene copy number variation in the human genome was previously unknown; and likewise, it is probably the case that the role of CNPs in common genetic diseases is underestimated. Many genetic diseases are caused by large-scale chromosomal duplications or deletions that alter the dosage of genes. Almost invariably, visible alterations of the genome result in severe neurological defects. This suggests that dosage-sensitive genes are very common in the genome and that the brain, due to its developmental complexity, is commonly affected by a chromosomal imbalance.Here, we present results of our research on the genetic causes of neurological disorders.

As an initial proof-of-principle, we present data on the analysis of patients with known cytogenetic abnormalities. ROMA scans were performed on three cases with cytogenetically detected chromosomal imbalances. Array data revealed that a del(13)(q14.3q21.2) was noncontiguous, a fact that was not apparent cytogenetically, but was later confirmed by FISH. A del(4)(q12q21.2) was found by ROMA to be 23 Mb in length, from 58.8 to 81.9 Mb on chromosome 4, in agreement with the cytogenetically assigned breakpoints. In a 3rd case, ROMA showed that an unbalanced "subtelomeric" rearrangement involved an unexpectedly large region of coding DNA including a 6 Mb deletion of 22q and an 8 Mb duplication of 16q. These results

serve both to validate the method and to illustrate the extraordinary precision with which we can now define chromosomal aberrations.

In addition, ROMA has proven to be a powerful tool for studies of complex genetic disease. Here we report the initial results of the analysis of copy number variation in autism. Our patient sample consisted of 180 familial cases (affected individuals form multiplex families) and 60 sporadic cases. We identified CNPs that were recurrent in autism and absent from a set of 91 controls, and some of these variants were located within autism-susceptiblity regions that others have identified by linkage analysis. Examination of the gene content of CNPs from autism patients revealed duplications and deletions affecting several genes that have been previously shown to be associated with autism and autism-related disorders. These data suggest that submicroscopic variation in gene copy number plays a role in common inherited psychiatric disorders such as autism.

Submicroscopic duplications and deletions: clinical implications of rearrangement mutations

James R. Lupski,
Baylor College of Medicine, Houston, Texas, U.S.A.,
and Wellcome Trust, Sanger Institute, Hinxton, Cambridge, U.K., jlupski@bcm.tmc.edu

For five decades since the seminal discoveries by Fred Sanger, that proteins have a specific structure, Linus Pauling that hemoglobin from patients with sickle cell anemia is molecularly distinct, and the elucidation of the chemical basis of heredity by Watson and Crick; the molecular basis of disease has been addressed in the context of how mutations effect the structure, function, or regulation of a gene or its protein product. Molecular medicine has functioned in the context of a genocentric world. During the last decade, however, it has become apparent that many disease traits are best explained not by how the information content of a single gene is changed, but rather on the basis of genomic alterations. Furthermore, it has become abundantly clear that architectural features of the human genome can result in susceptibility to DNA rearrangements that cause disease traits – such conditions have been referred to as genomic disorders.

In genomic disorders the clinical phenotype is a consequence of abnormal dosage or dysregulation of gene(s) due to rearrangements of the genome. Both inter- and intrachromosomal rearrangements are facilitated by the presence of low-copy repeats (LCRs) and result from nonallelic homologous recombinations (NAHR) between paralogous genomic segments. LCRs, also referred to as segmental duplications, span between 10 and several hundred Kb, share greater than 97-98% sequence identity, and provide the substrates for homologous recombination, thus predisposing the region to rearrangements. NAHR between directly oriented LCRs can result in deletion/duplication or gene conversion as alternate products of Holliday structure resolution; when LCRs in an inverted orientation are used as substrates, inversion of the flanked genomic segment occurs.

Analysis of the products of recombination at the junctions of recurrent rearrangements reveals "hotspots" for the crossovers and evidence consistent with double strand breaks as the initiating event. Thus, genome architecture consisting of LCRs including complex mosaic structures, both stimulates and mediates the rearrangement events. Non-recurrent rearrangements utilize both NAHR, with closely related repetitive sequences as substrates, and nonhomologous end joining (NHEJ) as causative mechanisms. For these latter rearrangement events, LCRs appear to stimulate, but not mediate, the genomic changes.

It remains to be determined to what extent genomic changes are responsible for disease traits, common traits (including behavioral traits), or perhaps sometimes represent benign polymorphic variation. It has become clear that high resolution analysis of the human genome has enabled detection of genome changes heretofore not observed due to technology limitations. Whereas conventional agarose gel electrophoresis allows resolution of changes of the genome up to 25-30 Kb in size, and cytogenetic banding techniques could resolve deletion rearrangements only greater than 2-5 Mb in size, alterations of the genome between > 30 Kb and < 5 Mb defied detection until pulsed-field gel electrophoresis (PFGE) and fluorescence *in situ* hybridization (FISH) became available to resolve changes in the human genome of such magnitude. These methods were limited to detection of specific genomic regions of interest and could not evaluate genomic rearrangements in a global way.

The availability of a "finished" human genome sequence and genomic microarrays have enabled approaches to resolve changes in the genome heretofore impossible to assess on a global genome scale; i.e. simultaneously examining the entire genome rather than discreet segments. Array Comparative Genome Hybridization (aCGH) is one powerful approach to high resolution analysis of the human genome. The CGH determines differences by comparisons to a reference "normal genome" whereas the array enables detection of such changes at essentially whatever resolution is desired, limited only by imagination and cost. Furthermore, the application of bioinformatics analyses to the finished human genome sequence and comparative genomic analysis enable information technology approaches to identify key architectural features associated with known recurrent rearrangements causing genomic disorders throughout the entire genome.

An increasing number of human diseases are recognized to result from recurrent DNA rearrangements involving unstable genomic regions. A combination of high resolution genomic analyses with informatics capabilities to examine individuals with well characterized phenotypic traits is a powerful approach to address the scientific question: To what extent are DNA rearrangements in the human genome responsible for human traits?

References:

1. Lupski, J.R. (1998). Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends in Genetics 14*:415-420.

2. Stankiewicz, P., and Lupski, J.R. (2002). Genome architecture, rearrangements, and genomic disorders. *Trends in Genetics 18*:74-82.

3. Shaw, C.J., and Lupski, J.R. (2004). Implications of human genome architecture for rearrangement based disorders: The genomic basis of disease. *Human Molecular Genetics 13*:R57-R64.

4. Lupski, J.R. (2004). Homologous recombination hotspots in the human genome – not all homologous sequences are equal. *Genome Biology 5*:242. http://genomebiology.com/2004/5/10/242.

5. Genomic Disorders – The Genomic Basis of Disease (2005). J.R. Lupski and P. Stankiewicz (Eds.) Humana Press, Totowa, N.J. In press.

Tiling-path genomic microarrays - applications in research and diagnostics

Jan P. Dumanski
Department of Genetics and Pathology, Rudbeck Laboratory,
Uppsala University, 751 85 Uppsala, Sweden
Email: jan.dumanski@genpat.uu.se, Home page:http://puffer.genpat.uu.se/

The development of comparative genomic hybridization (CGH) to metaphase spreads allowed genome-wide analysis of DNA copy number imbalances (1). This conceptually novel approach uses differentially labeled "test" and "reference" DNA, which are competitively hybridized to normal metaphase chromosomes on a glass slide. The ratio of the fluorescence intensities detected is indicative of the relative DNA copy number in test versus reference DNA. However, the main disadvantage of metaphase-CGH is its low resolution. Commonly mentioned figures in the literature are 5-10 Mb for deletions and 2 Mb for amplifications. This limitation was resolved with the advent of microarray-based comparative genomic hybridization (array-CGH). First described in 1997, array-CGH or matrix-CGH, paved the way for higher resolution detection of DNA copy number aberrations (2, 3). Array-CGH is based on the same principle(s) as metaphase-CGH, except that the targets are mapped genomic sequences instead of full chromosomes. This array-based approach introduced a new dimension in terms of resolution and specificity of analysis. Furthermore, this important improvement also allows a number of other assays to be performed (see below).

Applications of tiling-path arrays for human chromosome 22 and 1 in genetic and epigenetic analysis

In 2002, we constructed the first comprehensive microarray representing a human chromosome, namely chromosome 22, with the primary aim of performing reliable analysis of DNA copy number variation (4). This chromosome 22 array covers 34.7 Mb, representing 1.1% of the genome, with an average resolution of 75 kb. To demonstrate the utility of the array, we have applied it to profile acral melanoma, dermatofibrosarcoma, DiGeorge syndrome and neurofibromatosis type 2 (NF2) samples. We accurately diagnosed homozygous/hemizygous deletions, amplifications/gains, *IGLV/IGLC* locus instability, and the breakpoint(s) of an imbalanced translocation. We further identified the 14-3-3 eta isoform as a candidate tumor suppressor in glioblastoma. Significant methodological advances in array construction were also developed and validated. These include a strictly sequence defined, repeat-free, and non-redundant strategy for preparation of genomic arrays. This approach allows a dramatic increase in array resolution and analysis of any locus; disregarding common repeats, genomic clone availability and sequence redundancy. A broad spectrum of issues in medical research and diagnostics can be approached using our array. This well annotated and gene-rich autosome contains numerous uncharacterized disease genes. It is therefore crucial to associate these genes to specific 22q-related conditions and this array is being instrumental towards this goal. Since the publication of this chromosome 22 array, we have been applying it in the profiling of a large number of samples, including constitutional and/or tumor-derived DNA from: sporadic and familial ependymoma; sporadic breast-cancer; sporadic colon cancer; sporadic ovarian cancer;

sporadic or familial pheochromocytoma; sporadic Wilm's tumor; sporadic meningioma; sporadic glioblastoma; sporadic and NF2/schwannomatosis-associated schwannoma; familial schizophrenia as well as blood DNA from unrelated controls of different ethnic origin (5-13). This has already resulted in identification of several 22q loci, which are involved in the development/progression of these tumors or might predispose the patient to schwannoma, glioblastoma, breast- and ovarian-cancer as well as ependymoma development (6, 7, 9, 10).

Furthermore, we have also applied a human chromosome 1 tiling-path array constructed at the Wellcome Trust Sanger Institute UK, for the analysis of genetic and epigenetic changes in cancer-related-samples. One of the projects, aiming to analyze menigiomas, has already been completed (14). Meningiomas are common, usually benign neoplasms of the meninges lining of the central nervous system. Deletions of 1p have previously been established as important for the initiation and/or progression of meningioma. The rationale of this study was to characterize copy number imbalances of chromosome 1 in meningioma, using a full-coverage genomic microarray containing 2118 distinct measurement points. In total, 82 meningiomas were analyzed, making this the most detailed analysis of chromosome 1 in a comprehensive series of tumors. We detected a broad range of aberrations, such as deletions and/or gains of various sizes. Deletions were the predominant finding and ranged from monosomy to a 3.5 Mb terminal 1p homozygous deletion. Although multiple aberrations were observed across chromosome 1, every meningioma in which imbalances were detected harbored 1p deletions. Tumor heterogeneity was also observed in three recurrent meningiomas, which most likely reflects a progressive loss of chromosomal segments at different stages of tumor development. The distribution of aberrations supports the existence of at least four candidate loci on chromosome 1, which are important for meningioma tumorigenesis. In one of these regions, our results already allow the analysis of a limited number of candidate genes, which is ongoing in a follow up study. Furthermore, in a large series of cases we observed an association between deletion breakpoints and the presence of segmental duplications, which suggests their role in the generation of these tumor-specific aberrations. As 1p is the site of the genome most frequently affected by tumor-specific aberrations, our results indicate loci of general importance for cancer development and progression.

The term epigenetics can be defined as hereditable or acquired changes in gene expression, which are not associated with changes in DNA sequence. The best-known epigenetic modification is DNA methylation, which tags cytosine in CpG dinucleotides with a methyl group. DNA methylation of CpG islands within the promoter and/or the first exon-intron region of genes, is generally associated with silencing of gene expression. Chemical modification of histones can also function as an epigenetic mechanism and genes associated with acetylated histones are usually turned on. DNA microarrays can be applied to map the location of chemical modifications of chromatin and in the identification of chromosomal targets of proteins involved in DNA binding or chromatin remodeling (15). The "ChIP-to-chip" technique, which combines Chromatin Immuno-Precipitation (ChIP) with hybridization to DNA arrays (the chip), has proven to be a powerful method to elucidate the interaction sites of DNA-binding proteins across the genome (16, 17). Aberrant DNA methylation within CpG islands is among the first and most common alteration leading to silencing or over-expression of genes in cancer (18, 19) and are therefore potentially promising tools for early diagnostics (20). Conventional methods used to determine methylation status limit the evaluation to only a single/few selected genes. Incorporation of epigenetic analysis methods into microarray technology can yield quantitative and qualitative information of DNA methylation across the genome (15, 21). We are currently analyzing CpG island methylation changes on chromosome 1 in samples derived from families

affected with ICF syndrome, normal fully differentiated adult human tissues and colon cancer cell-lines which are deficient in DNMT1 and DNMT3b methylases (manuscript in preparation). For that DNA is cleaved with a methyl-sensitive restriction endonuclease, size fractionated (21) and hybridized to the chromosome 1 microarray. The results obtained so far indicate that tiling-path BAC-microarrays clearly allow discriminating chromosomal domains that are subject to differential methylation.

Array-CGH in diagnostics

Genetic aberrations at the DNA copy number level such as deletions, gains/amplifications and unbalanced translocations are among the main pathogenetic mechanisms behind human genetic disorders (http://www.ncbi.nlm.nih.gov/omim/). Analysis of chromosomal aberrations is especially important in cancer, where amplification of oncogenes and/or deletion of tumor suppressor genes are important events in the multi-step process of cancer development (22). Therefore, rapid and accurate identification of such genetic imbalances could improve diagnostics, which may lead to better decision-making regarding the choice of available treatments. However, in spite of the importance of genetic imbalances as disease-causing factors, the routine approaches for detection of such imbalances are underdeveloped. This need for improved methodology for DNA copy number analysis is especially apparent when compared to a variety of methods available for accurate detection of point mutations. Arrays focused on selected segments of the genome, known to be frequently associated with specific diseases have been successfully applied (13, 23). One such example is neurofibromatosis type 2 (NF2) (MIM no. 101000), which is an autosomal dominant disorder caused by the bi-allelic inactivation of the *NF2* gene from 22q12. Affected individuals usually develop bilateral schwannomas of the eighth cranial nerve but can also be affected with other tumors. Inactivating deletions in the *NF2* gene occur in the range of 20-30% of NF2 patients. A high-resolution diagnostic microarray for profiling DNA copy number alterations in the *NF2* gene has been developed (12, 24).

We have also recently completed the construction and application of a neurofibromatosis type 1 (NF1) locus-specific diagnostic microarray, for detection of *NF1* gene micro-deletions(25). Segmental duplications flanking the neurofibromatosis type 1 (*NF1*) gene locus on 17q11 mediate the majority of gene deletions in NF1 patients. However, the large size of the gene and the complexity of the locus architecture pose difficulties in the deletion analysis. This first *NF1* locus-specific microarray covers 2.24 Mb of 17q11 and is constructed using the above mentioned PCR-based, non-redundant approach for array design. The average resolution of analysis for the array is ~12 kb per measurement point with an increased average resolution of 6.4 kb for the *NF1* gene. We performed a comprehensive array-CGH analysis of 161 NF1 constitutional or tumor-derived samples and identified heterozygous deletions of various sizes in 39 cases. The typical 1.44 Mb deletion was identified in 26 cases, whereas 13 samples showed atypical deletion profiles. The size of the atypical deletions, contained within the segment covered by the array, ranged from 6 kb to 1.6 Mb and their breakpoints could be accurately determined. Moreover, 10 atypical deletions were observed to share a common breakpoint either on the proximal or distal end of the deletion. The deletions identified by array-CGH were independently confirmed using multiplex ligation-dependent probe amplification (MLPA). Bioinformatic analysis of the entire locus identified 33 segmental duplications. We show that at least one of these segmental duplications, which borders the proximal breakpoint located within the *NF1* intron-1 in five atypical deletions, might represent a novel hot-spot for deletions. Our array constitutes a novel and reliable tool for significant improvements of diagnostics for NF1, which is a very common disorder (MIM no. +162200).

Summary

Genomic tiling-path high-resolution genomic microarrays provide a versatile tool for assessment of gene copy number changes in health and disease, epigenetic modifications, transcription factor binding sites, replication timing, annotation of new transcriptional units as well as mapping of balanced and unbalanced translocations. It is likely that the widespread use of these tools will make a major impact on medical research and diagnostics in the coming years (26).

The study of human genetic variation at the DNA level constitutes a major challenge which has recently received considerable attention. However, the dominating type of variation explored so far in the genome has been at the level of single nucleotide polymorphisms (SNPs), overshadowing the issue of copy number polymorphisms (gains and deletions, CNPs). The current approach to study genetic variation may thus be viewed as strongly biased, in the sense that the identification of genome-wide large scale CNPs is virtually untouched, when compared to detailed analyses of SNPs in their millions. It is only when parallel analyses of both types of variation are performed, that we will obtain a more complete picture of our genetic diversity.

As opposed to SNPs, the analysis of CNPs is considerably more difficult, due to the complexity of variation and costs involved in their detailed assessment. It is not only the presence or absence of the CNP which matters, but the type (deletion or gain) and size of the copy number variation must also be considered. Currently, the approaches for analysis of copy number variation include the use of genomic clones (BACs, PACs and cosmids), cDNA clones, PCR products as well as spotted or *in situ* synthesized oligonucleotides. The comparability of the results generated from these different platforms is currently the major obstacle in the field and contributes to confusion in data interpretation between different reports. Furthermore, there is a lack of common uniform criteria for the quality assessment of published array-CGH data and a standard similar to the expression array MIAME (minimum information about microarray experiment; http://www.mged.org) criteria should be introduced (27).

References

1. Kallioniemi, A., Kallioniemi, O. P., Sudar, D., Rutovitz, D., Gray, J. W., Waldman, F., and Pinkel, D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science, 258: 818-821, 1992.
2. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T., and Lichter, P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer, 20: 399-407, 1997.
3. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W. L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W., and Albertson, D. G. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet, 20: 207-211, 1998.
4. Buckley, P. G., Mantripragada, K. K., Benetkiewicz, M., Tapia-Páez, I., Diaz de Ståhl, T., Rosenquist, M., Ali, H., Jarbo, C., de Bustos, C., Hirvelä, C., Sinder Wilén, B., Fransson, I., Thyr, C., Johnsson, B.-I., Bruder, C., Menzel, U., Hergersberg, M., Mandahl, N., Blennow, E., Wedell, A., Beare, D., Collins, J., Dunham, I., Albertson, D., Pinkel, D., Bastian, B., Faruqi, A., Lasken, R., Ichimura, K., Collins, V., and Dumanski, J. P. A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications. Human Molecular Genetics, 11: 3221-3229, 2002.

5. Jarbo, C., Buckley, P., Piotrowski, A., Mantripragada, K., Benetkiewicz, M., Diaz de Ståhl, T., Langford, C., Gregory, S., Dralle, H., Gimm, O., Bäckdahl, M., Geli, J., Larsson, C., Westin, G., Åkerström, G., and Dumanski, J. Detailed assessment of chromosome 22 aberrations in sporadic pheochromocytoma using array-CGH. International Journal of Cancer, accepted, pending revisions, 2005.
6. Díaz de Ståhl, T., Hartmann, C., de Bustos, C., Piotrowski, A., Benetkiewicz, M., Mantripragada, K., Tykwinski, T., von Deimling, A., and Dumanski, J. Chromosome 22 tiling path array-CGH analysis identifies germline- and tumor-specific aberrations in patients with glioblastoma multiforme. Genes, chromosomes & cancer, in press, 2005.
7. Diaz de Ståhl, T., Hansson, C., de Bustos, C., Mantripragada, K., Piotrowski, A., Benetkiewicz, M., Wiklund, L., Mathiesen, T., Nyberg, G., Collins, V., Evans, D., Ichimura, K., and Dumanski, J. High-resolution array-CGH profiling of germline and tumor- specific copy number alterations on chromosome 22 in patients affected with schwannomas. Human Genetics, in press, 2005.
8. Buckley, P., Mantripragada, K., Díaz de Ståhl, T., Piotrowski, A., Hansson, C., Kiss, H., Vetrie, D., Ernberg, I., Nordenskjöld, M., Bolund, L., Sainio, M., Rouleau, G., Niimura, M., Evans, D., Grigelionis, G., Menzel, U., and Dumanski, J. Evidence for genetic aberrations on chromosome 22 outside the NF2 locus in schwannomatosis and neurofibromatosis type 2. submitted, 2005.
9. Benetkiewicz, M., Wang, Y., Schaner, M., Wang, P., Mantripragada, K. K., Buckley, P. G., Kristensen, G., Borresen-Dale, A. L., and Dumanski, J. P. High-resolution gene copy number and expression profiling of human chromosome 22 in ovarian carcinomas. Genes Chromosomes Cancer, 42: 228-237, 2005.
10. Ammerlaan, A., de Bustos, C., Ararou, A., Buckley, P., Mantripragada, K., Verstegen, M., Hulsebos, T., and Dumanski, J. Localization of a Putative Low-Penetrance Ependymoma Susceptibility Locus to 22q11 Using a Chromosome 22 Tiling-path Genomic Microarray. Genes, chromosomes & cancer, in press, 2005.
11. Mantripragada, K. K., Tapia-Páez, I., Blennow, E., Nilsson, P., Wedell, A., and Dumanski, J. DNA copy-number analysis of the 22q11 deletion-syndrome region using array-CGH with genomic and PCR-based targets. Int J Mol Med, 13: 273-279, 2004.
12. Mantripragada, K. K., Buckley, P. G., Jarbo, C., Menzel, U., and Dumanski, J. P. Development of NF2 gene specific, strictly sequence defined diagnostic microarray for deletion detection. J Mol Med, 81: 443-451, 2003.
13. Mantripragada, K. K., Buckley, P. G., Benetkiewicz, M., De Bustos, C., Hirvela, C., Jarbo, C., Bruder, C. E., Wensman, H., Mathiesen, T., Nyberg, G., Papi, L., Collins, V. P., Ichimura, K., Evans, G., and Dumanski, J. P. High-resolution profiling of an 11 Mb segment of human chromosome 22 in sporadic schwannoma using array-CGH. Int J Oncol, 22: 615-622, 2003.
14. Buckley, P. G., Jarbo, C., Menzel, U., Mathiesen, T., Scott, C., Gregory, S. G., Langford, C. F., and Dumanski, J. P. Comprehensive DNA copy number profiling of meningioma using a chromosome 1 tiling-path microarray identifies novel candidate tumor suppressor loci. Cancer Research, 65: 2653-2661, 2005.
15. Pollack, J. R. and Iyer, V. R. Characterizing the physical genome. Nat Genet, 32 Suppl: 515-521, 2002.

16. Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P., and Young, R. A. Genome-wide location and function of DNA binding proteins. Science, 290: 2306-2309, 2000.
17. Iyer, V. R., Horak, C. E., Scafe, C. S., Botstein, D., Snyder, M., and Brown, P. O. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature, 409: 533-538, 2001.
18. Jones, P. A. and Laird, P. W. Cancer epigenetics comes of age. Nat Genet, 21: 163-167, 1999.
19. Ehrlich, M., Jiang, G., Fiala, E., Dome, J. S., Yu, M. C., Long, T. I., Youn, B., Sohn, O. S., Widschwendter, M., Tomlinson, G. E., Chintagumpala, M., Champagne, M., Parham, D., Liang, G., Malik, K., and Laird, P. W. Hypomethylation and hypermethylation of DNA in Wilms tumors. Oncogene, 21: 6694-6702, 2002.
20. Laird, P. W. The power and the promise of DNA methylation markers. Nat Rev Cancer, 3: 253-266, 2003.
21. Tompa, R., McCallum, C. M., Delrow, J., Henikoff, J. G., van Steensel, B., and Henikoff, S. Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. Curr Biol, 12: 65-68, 2002.
22. Kinzler, K. W. and Vogelstein, B. Lessons from hereditary colorectal cancer. Cell, 87: 159-170, 1996.
23. Bruder, C. E., Hirvela, C., Tapia-Paez, I., Fransson, I., Segraves, R., Hamilton, G., Zhang, X. X., Evans, D. G., Wallace, A. J., Baser, M. E., Zucman-Rossi, J., Hergersberg, M., Boltshauser, E., Papi, L., Rouleau, G. A., Poptodorov, G., Jordanova, A., Rask-Andersen, H., Kluwe, L., Mautner, V., Sainio, M., Hung, G., Mathiesen, T., Moller, C., Pulst, S. M., Harder, H., Heiberg, A., Honda, M., Niimura, M., Sahlen, S., Blennow, E., Albertson, D. G., Pinkel, D., and Dumanski, J. P. High resolution deletion analysis of constitutional DNA from neurofibromatosis type 2 (NF2) patients using microarray-CGH. Hum Mol Genet, 10: 271-282, 2001.
24. Dhami, P., Coffey, A. J., Abbs, S., Vermeesch, J. R., Dumanski, J. P., Woodward, K. J., Andrews, R. M., Langford, C., and Vetrie, D. Exon Array-CGH: Detection of Copy-Number Changes at the Resolution of Individual Exons in the Human Genome. American journal of human genetics, 76: 750-762, 2005.
25. Mantripragada, K., Thuresson, A., Piotrowski, A., Díaz de Ståhl, T., Menzel, U., Grigelionis, G., Bolund, L., Mautner, V., Legius, E., Nordling, M., Vetrie, D., Dahl, N., Messiaen, L., Upadhyaya, M., Bruder, C., and Dumanski, J. Identification of deletion breakpoints bordering novel segmental duplications by comprehensive array-CGH analysis of NF1 samples using an NF1 locus microarray. Journal of Medical Genetics, accepted, pending revisions, 2005.
26. Mantripragada, K. K., Buckley, P. G., Diaz de Ståhl, T., and Dumanski, J. P. Genomic microarrays in the spotlight. Trends Genet, 20: 87-94, 2004.
27. Buckley, P., Mantripragada, K., Piotrowski, A., Díaz de Ståhl, T., and Dumanski, J. Copy number polymorphisms: mining the tip of an iceberg. Trends in Genetics, in press, 2005.

Genetic Analysis of Syndrome X on the Pacific Island of Kosrae

Jeffrey M. Friedman,
The Rockefeller University, New York, USA

The molecular basis of Syndrome X is being explored in genetic studies of obesity, diabetes, hypertension and hyperlipidemia among the inhabitants of the Pacific Island of Kosrae in Micronesia. These studies are being conducted in collaboration with the Kosrae Department of Health. The Kosraen population is extensively admixed between Caucasians and Micronesians and the population has a high incidence of obesity and diabetes.

In 1994, a comprehensive medical work up was performed on the entire adult population, approximately 2100 individuals. Each of these individuals has been placed into a single extended pedigree. Genetic analysis of DNA from these individuals has localized several loci that appear to contribute to the development of obesity and associated conditions in this population.

In 2002, a second study was concluded in which 2173 individuals were screened. This screening included a number of studies that were not performed in 1994 including direct measures of body composition by impedance measures and a large number of relevant blood tests. To date a total of 3212 individuals have been studied and longitudinal data is available for 1043 of the participants. A genome scan using 100K Affymetrix SNP microarrays is being generated for all 3212 participants. An analysis of the data thus far reveals significantly longer segments of LD and reduced haplotype diversity both of which may facilitate genetic analyses. Other studies have identified Mendelian loci on Kosrae with high allele frequencies suggesting that homozygous mapping might prove useful for identifying possible single gene disorders for one or more traits. These approaches will be used to map and eventually identify genes responsible for obesity , diabetes, hypertension and hyperlipidemia in this population.

What can progeroid syndrome tell us about normal human ageing?

Richard G.A. Faragher
SPARC Programme Director, School of Pharmacy and Biomolecular Science
University of Brighton, Brighton, BN2 4GJ
Tel: +44 1273 642124 Fax: +44 1273 679333

The ageing of the European population is an urgent social-economic challenge. By 2050 approximately 40% of the population of Europe will be over the age of 60 (currently less than 15% of persons fall into this category). Whilst increasing numbers of long-lived people are a testament to Europe's social stability and scientific advances ill health, frailty and dependency remain far too common among our elderly. Thus, whilst ageing is associated with an increased chance of death it is this increase of morbidity (and the resultant loss of quality of life) which is the key challenge of the ageing population.

The costs of the ageing process are substantial. It has been estimated that a 1% reduction in the rate of morbidity within one member state alone (the United Kingdom) would result in a saving of 75 billion NKr per annum. In order to achieve the compression of morbidity it is essential to understand the biological basis of the normal ageing process.

As an evolutionary side effect of millions of years of selection for reproductive success, the genetic basis of aging is potentially extremely broad. Perhaps as much as 7% of the total genome may play a role in determining the lifespan of humans. With such a highly polygenic system, to suggest that any single mechanism or simple network of mechanisms is responsible for the aging of a whole organism is rather naive. By the same logic the identification of candidate genes involved in successful aging through the study of normal centenarian "survivor" populations, whilst potentially feasible, is proving an extremely complicated task. An alternative approach, first articulated as a formal concept by George Martin, is the study of heritable genetic diseases which mimic some, but not all, the features of the aging process in order to gain insights into how the aging process functions in normal individuals. The study of such "progeroid" syndromes has the advantage that only a single gene is usually involved in each case. This renders hypotheses easier to frame and test and makes the manipulation of the candidate allele possible in a way denied to workers studying highly polygenic traits. The disadvantage of studying progeroid syndromes is that they are essentially phenocopies of normal aging rather than the genuine article. Any observations made using them must thus be evaluated within the context of theories designed to explain how normal aging operates.

The best characterised of these disorders is Werner's syndrome [1]. This is an autosomal recessive genetic disorder is in which affected individuals prematurely display many of the features associated with old age. Werner's syndrome patients show cardiovascular disease (both atherosclerosis, and arteriosclerosis), greying of the hair, cancer, bilateral cataracts, osteoporosis, type II diabetes and thymic atrophy. Death occurs at an average age of 47. It is this unique clinical presentation (rather than any *a priori* knowledge of the biological mechanism causing it) that renders the disease so interesting to gerontologists.

Classical Werner's syndrome results from loss of function mutations in a gene coding for a member of the recQ helicase family (designated *wrn*) [2]. Lack of *wrn* causes an elevated frequency of DNA replication fork stalling. This produces hypersensitivity to some types of DNA damaging agents, a distinct mutator phenotype characterised by large DNA deletions [3] and premature replicative senescence in a number of cell types [4]. In order for Werner's syndrome to be a useful model of ageing it is important to understand whether the degenerative pathology observed in the patients primarily results from the increased degree of genomic instability or increased rates of replicative senescence that are seen at the cellular level. In addition two distinct strains of Werner's syndrome transgenic mice have now been generated, one of these has no pathology [5]whilst the other recapitulates virtually all the clinical features of the disease [6].

It is my view that the data available to us from the study of Werner's syndrome have been extremely valuable in advancing our understanding of the normal aging process in two distinct ways. Firstly, consideration of the disease has been useful in a methodological sense because it has driven the formulation of extremely precise hypotheses that are amenable to clear refutation and are thus (in Karl Popper's sense of the term) extremely powerful. Secondly, data arising from the study of Werner's syndrome have been practically informative in distinguishing between (or refining) hypotheses already widespread within gerontology. In my view, the existing data from Werner's syndrome:

i. Are consistent with the hypothesis that the replicative senescence of human cells is the primary causal mechanism of the disease pathology.

ii. Provide visible limits to the replicative senescence hypothesis that emphasize the multicausal nature of the aging process.

iii. Stress the context dependency of senescence mechanisms both between different tissues in the same organism and between the same tissue in different species. This latter element has implications not simply outside the study of the disease, but outside gerontology as well.

References

[1] Kipling D, Ostler E.L., Davis T. & Faragher R.G.A. (2004) What can progeroid syndromes tell us about human aging? *Science* **305**:1426-31

[2] Yu CE, Oshima J, Ying-Hui F, Wijsman EM, Hisami F, Alisch R, Matthews S, Nakura J, Miki T, Ouais S , Mulligan J, Martin GM, Schellenburg GD. Positional cloning of the Werner's syndrome gene. *Science* 1996; **272:** 258-262

[3] Fukuchi K, Martin, GM, Monnat RJ. Mutator phenotype of Werner syndrome is characterised by extensive deletions. *Proc. Natl. Acad. Sci. USA* 1989; **86:** 5893-5897.

[4] Faragher RGA, Kill IR, Tannock TCA, Hunter JAA & Shall S. Is the Werner's Syndrome Gene a "Clock" Gene which counts Cell Division? *Proc. Natl. Acad. Sci*. 1993;**90:** 12030-12034.

[5] Lebel M, Leder P. A deletion within the murine Werner syndrome helicase induces sensitivity to inhibitors of topoisomerase and loss of cellular proliferative capacity. *Proc Natl Acad Sci U S A.* 1998; **95**:13097-102

[6] Chang S, Multani AS, Cabrera NG, Naylor ML, Laud P, Lombard D, Pathak S, Guarente L, DePinho RA. Essential role of limiting telomeres in the pathogenesis of Werner syndrome. *Nat Genet*. 2004 **36**:877-82.

Genetics of Natural Variation in Human Gene Expression

Vivian G. Cheung,
University of Pennsylvania, Philadelphia, Pennsylvania, USA.

Gene expression plays an important role in defining cellular phenotype and function. Natural variation in human gene expression is the focus of this study. The goal is to identify the extent of variation in transcript level of genes among normal individuals and to determine whether there is an inherited basis to this variation.

First, to assess the variation in gene expression, we used microarrays to measure the expression level of ~ 8500 genes in lymphoblastoid cells from 50 unrelated individuals[1,2]. All experiments were performed in duplicate. For each gene, we calculated the variance ratio, the variance of expression level among individuals divided by variance among array replicates. This allows us to identify those that are the most variable among individuals relative to measurement noise. Among the 8500 genes on the microarrays, ~ 3800 genes were expressed in at least 10 of the 50 individuals studied and these expressed genes were the focus of our analysis. Among them, the variance ratios ranged from 0.2 to 48.9 (mean = 2.6, median = 1.6). The array replicates are highly similar; the variance ratios for 3,554 genes (> 90%) are greater than 1. For our study, we focused on these genes where the inter-individual variation is greater than technical variation. Some of these genes are highly variable. For the 50 most variable genes, the range of their expression levels is at least 15-fold.

Next, we determined if there is a genetic basis to this variation. For 15 of the variable genes, we measured their expression level in individuals with different degree of relatedness. We examined their expression in monozygotic twins, and among siblings within families and in unrelated individuals. For each of the genes, we found the extent of variation was highest among the unrelated individuals and lowest among members of twin pairs. The variability increases as the degree of relatedness decreases suggesting a genetic component to the variation in gene expression. We followed up this finding by a more standard analysis and calculated the narrow-sense heritability. We measured the expression level of genes in 50 parent-child trios. For each gene, we performed regression of the expression level of the offspring on mid-parent value as an estimate of heritability. This regression analysis was performed on the 3,554 genes that are expressed in lymphoblastoid cells. About half of these values were positive suggesting some genetic component. The regression coefficient is >0.5 for 94 genes and >0.75 for 9 genes. These estimates are at best crude indicators of heritability. But along with the comparison of variance ratios among individuals with different degree of relatedness, these regression estimates suggest that there is a genetic component to the variation in gene expression.

To extend these results, we performed genetic analyses to identify the determinants of gene expression variation[3]. To localize the genetic determinants of these quantitative phenotypes, we measured the expression level of genes in lymphoblastoid cells from members of 14 Centre d'Etude du Polymorphisme Humain (CEPH) Utah pedigrees and obtained SNP genotypes of the same individuals from the genotype database of the SNP consortium. We carried out genome-wide linkage analysis for 3,554 gene expression phenotypes using the computer program S.A.G.E.. The analysis gives the strength of the evidence at each map position as the value of a t-statistic with a corresponding pointwise P-value. Among the phenotypes examined, 142 have evidence of linkage beyond a point-wise P-value of 4.3×10^{-7}, t>5 (genome-wide P<0.001). At a

lower threshold of point-wise $P<3.7 \times 10^{-5}$ ($t>4$), 984 phenotypes exceeded this threshold. This is far greater than the 178 false positives expected by chance.

The regions that are linked to the expression levels are likely transcriptional regulatory sites. Among the 142 phenotypes with the strongest evidence of linkage, 27 have only a cis-acting regulator, 110 have only a trans-acting regulator and 5 have more than one regulator. To allow for imprecision of linkage and long-range regulators, we defined cis-regulation as those that mapped within 5 megabase of the target gene.

In the genome, there are sites that appear to be master regulatory region of transcription. These genomic regions contain more transcriptional regulators than expected by chance. Two such examples are 5 megabase regions on chromosome 14q32 and 20q13 where linkages for 31 phenotypes and 25 phenotypes were mapped respectively.

In this presentation, I will discuss results described above and also present data from association study to replicate and narrow candidate regulatory regions identified by the linkage analysis.

1. Cheung, V. G. et al. Natural variation in human gene expression assessed in lymphoblastoid cells. *Nat Genet* **33**, 422-25 (2003).
2. Cheung, V. G. et al. Genetics of quantitative variation in human gene expression. *Cold Spring Harbor Symp. Quant. Biol.* **68**, 403-407 (2003).
3. Morley, M. et al. Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743-7 (2004).

SNPS in putative regulatory loci controlling gene expression in cancer

Vessela N. Kristensen,
Instiute for Cancer Research,
The Norwegian Radium Hospital, Norway

Given the increasing clinical importance of microarray expression classification of breast tumours and the different biology it may reveal[1], identifying an associated SNP profile may be of considerable value for pharmacogenetics, early diagnostics and cancer prevention. Studying the promoter composition of the genes that strongly predict the patient subgroups we observed clear separation of the gene clusters based solely on their promoter composition, making feasible the hypothesis that SNPs in the regulatory regions of genes that create or abrogate transcription binding sites have the potential to influence the expression profiles. Morley et al.[2] reported linkage analysis of expression levels of 3554 genes and 2500 SNPs in 14 CEPH families (retrieved from http://snp.cshl.org/), and found significant evidence for the existence of regulation hot spots, suggesting both *cis* and *trans* regulatory effects. We report similar observations from a study with a different design, performing actual genotyping of 49 unrelated breast cancer patients, whose tumours have previously been analyzed by genome-wide expression microarrays leading to a robust tumour classification with strong prognostic impact[3]. These patients were a part of a pharmacogenetic study of 193 patients who had received radiation therapy or chemotherapy. A high throughput solid phase, array based method using primer extension chemistry has been used to perform the genotyping (GenomeLab™ SNPstream genotyping system, Beckman Coulter). A total of 583 SNPs in 203 selected genes (1-19 SNPs/gene) were genotyped and tumour genome-wide expression was studied in 49 patients. Association in both *cis* and *trans* was detected for SNPs in 42 genes. SNP-expression associations with the top 0.25% best p-values ($9.81E\text{-}06<p<0.001$) revealed regulatory SNPs in 115 genes *in trans*. Therefore, the *trans* interactions observed here suggest both "positional" and "functional" explanations. The "positional" scenario may involve either linkage disequilibrium or epigenetic events such as common domains of relaxed chromatin structure along the chromosomes. The "functional scenario" may involve interactions of the kind between receptors and their ligands, transcription factors and genes under their control, the genes for which do not have to reside in vicinity to each other. Of that kind we observed a strong association between SNPs in several growth factors and the expression of their receptors like *EGF*/*EGFR, IL1*/*IL1R*, and *TGFB2/TGFBR1,R2,R3.* In fact in this study we use the genotypes from a given pre-defined pathway (the ROS metabolizing and signaling pathway) enriched for GO terms like ATP binding, phosphate metabolism, phosphorylation and tyrosine kinase activity to fish for associations from a whole genome pull of mRNA transcripts. If we inversely look at the GO terms of the mRNA that we find associated to these SNPs, we find often the same GO terms as for the candidate genes, suggesting that the observed associations are within the same functional pathway. This raises the question how a SNP can affect the expression level in an interacting gene without always apparently affecting the expression of its own gene. In a whole-genome analysis a certain cascade effect could be expected– only the most noticeable fluctuation in expression and the most robust interactions will remain "visible", leaving the possibility that there are in fact more *cis* interactions in nature requiring more sensitive functional methods.

SNP signals associated to gene expression were observed in lymphoblastoid cell lines from healthy individuals[2] and, in the present study, in blood DNA from breast cancer patients associated to expression in the tumor tissue. We may expect different, stronger signals in our study, admitting the existence of strong SNPs or expression susceptibility pattern associated with breast cancer per se. Indeed, we recognized important regulators of whole pathways such as *NFKB1*, *EGF* and *FGF2* among the genes in which SNPs have impact on mRNA expression. Still more profound functional studies are necessary to prove the causal relationship and to grant these SNPs "regulators" status. Taken together these data suggest that the observed SNP-expression associations do exist and are observable even in a small set of unrelated individuals. A given expression profile of the tumour may be potentially associated and predicted by the genotype of the patient.

[1] Perou. C.M. *et al. Nature.* **406,** 747-452 (2000).
[2] Morley, M. *Et al. Nature.* **430**, 743-747 (2004).
[3] Sorlie, T. *et al. Proc Natl Acad Sci USA*. **98,** 10869-10874 (2001).

The Estonian Genome Project

Andres Metspalu,
Department of Biotechnology, IMCB, University of Tartu and Estonian Biocenter, MDC of the Joint Laboratories of University of Tartu Clinicum and Estonian Genome Project Foundation, Tartu, Estonia. e-mail: andres@ebc.ee

In order to apply the personalized medicine we have to be more personal than just only the individual SNP map. We have to know the patient's other variables as well, such as environmental factors, history of exposure, health status, lifestyle, diet and other classical epidemiological data. And we have to keep strongly in mind the confounding, bias and reproducibility when designing and performing these studies. This brings objectively to the large, prospective population-based cohorts – biobanks - as one of the last missing resource after the HapMap and genotyping technology together with the bioinformatics in order to move towards the personalized medicine. This very fusion of different disciplines brings together high-tech genetics and large-scale epidemiology and will be the discovery field for the next few decades. Ideally, we all should be in the biobank and by implementing „e-health"concept widely enough this will eventually happen.

The UK Biobank (www.ukbiobank.ac.uk) has started a pilot study after careful preparations and the main project should start in 2006. The Estonian Genome Project (www.geenivaramu.ee) has 1% of the adult population (10 300 gene donors) in the database so far and the first research projects have started. There are more biobank projects in different stages with different scope, design and goals. The international consortium „Public Population Project in Genomics – P3G" (www.p3gconsortium.org) has been organized in order to promote collaboration between the different biobanks. One question is how similar or different are the European populations compared to the CEPH data from the HapMap project. It is seems to be clear now that if we use HapMap tag SNPs we can get more that 90% of the LD structure from any European population. Next 3-5 years and tens of the whole genome association studies with 250 to 500K SNPs will have unprecedented impact to human genetic medicine.

Challenges for epidemiology in the post-genome era – towards a less reductionist approach?

Eiliv Lund
University of Tromsø, Norway

The description of the human genome in 2001 was said to change medical research. For epidemiology the change is still under way dependant on changes in research design and technology improvements.

Analytical epidemiological methods for observational studies have two major design options; case-control and cohort (prospective) study design. According to the methods for collection of biological samples, mainly blood, several new analytical strategies have been possible.

As an example of new design options and analytical approaches the Norwegian Women and Cancer postgenome cohort study, NOWAC, will be introduced.

Traditionally a cohort study consists of questionnaire information on lifestyle and a biobank with plasma/serum and "buffy coat" or whole blood for extraction of DNA. In NOWAC we have added two distinct features. First, we collect blood samples with a later for preservation of RNA. Secondly, we attempt to collect biopsy material from beast cancer arising among the cohort members.

This design has given us new analytical options. We can have a *case-case design* for looking at the expression patterns among cases of breast cancer comparing expression pattern at start of follow-up with expression pattern in peripheral blood at time of diagnosis, and expression pattern of the tumours. We can further have a *combined prospective and case-control design or a quadratic design* by comparing expression pattern of cases with expression patterns of controls drawn within the cohort, eventually adjusting for lifestyle factors.

Lastly, we can try to use the patterns of expression for mechanistic studies of metabolism or biochemistry in a situation where we also deal with exposure and disease outcome. This would merge epidemiology, genetic and biochemical research into a complicated structure leaving the reductionist approach to cancer that each discipline has in itself.

Genetic epidemiology of BRCA1 mutations in Norway

Pål Møller,
Section of Genetic Counselling, Department for Cancer Genetics,
The Norwegian Radium Hospital, Oslo, Norway

Any scientific description is based on an understanding of what the problem to describe is, and inside which context the problem is addressed. Furthermore, within this frame, any description is based on an *a priori* understanding of within which framework of understanding the solution is to be found. This is usually referred to as scientific paradigms. It is not a question of applying paradigms or not, the question is to understand which paradigms you base your reasoning on at any time. These paradigms are often not mentioned, they are often socially determined, and you are soon to become an outlaw in the scientific society if you dear to address the paradigms. History is full of well known examples of shift in scientific paradigms, the social problems around shifting to the concept of the heliocentric understanding of the universe (Galileo) and the description of evolution by selection (Darwin) is common knowledge. Man is, however, still a tribal animal, and within the scientific tribe the current paradigms are seldom addressed because it is still associated with social insecurity to leave the herd and base your reasoning on any different paradigm than those currently applied.

An inherent property of most paradigms, is that they are self-fullfilling when entered into the classical scientific pathway of testing a hypothesis deduced from the current paradigm. This is so, because most (all?) current paradigms have some element of truth, and most (all?) hypothesis derived from them will have some element of truth. In sum, there will be an association to detect, the p value will – pending sample size – be less than 0.05, and the hypothesis will be demonstrated to be 'true'. The simultaneous fact is, however, that there may be other paradigms leading to other hypotheses that may be equally or more 'true' (lower p value). In this way, if all hypotheses tested are derived from current paradigms, you may never see that current paradigms are insufficient as driving forces to gain knew knowledge – even if they are.

What then, is the relevance to the above arguments with respect to genetic epidemiology of BRCA1 mutations in Norway?

The relevance is that most genetic epidemiologists apply the paradigm of Hardy-Weinberg equilibrium, and regards deviations from the paradigm as statistical confounders in the analyses. In society, this paradigm has been merged with Darwinism to become something like a religion: Living organisms no longer has a genetically determined behavioural pattern aquired by selection of the fittest, - thy now in all TV educational programmes have behavioural patterns with an intent: They do so because they want the effect achieved by doing so. More so, to protect the planet and our common environment, we must not disturb the 'balance' in the nature. The paradigm that nature is in balance has obviously merged with religious conceptions of using selection as a tool to arrive at the balance of today. The conflict between Darwinism and the religion of my society, has in my lifetime been solved by adopting Darwinism as a tool used in the process of making life as we know it. Other mechanisms – like earthcrakes, tsunamis, rain-floods – have been subjected to concepts of negative forces disturbing the balance and somehow in conflict with the making of earth and life – they all disturb the balance which is good and given and demonstrated to exist.

In contrast to current dominating paradigms, today's environment – including genetic epidemiology of man – is basically influenced of factors destroying balance. Actually, possibly no population at all is meeting the concept of Hardy-Weinberg equilibrium – and that concept actually relates to sexual breeding only. Most forms of life may use other mechanisms. The tulip is my favourite: It clones itself beneath the surface in any weather, but sticks it's head up and have sex when the sun is shining.

Again, what is the relevance to the above arguments with respect to genetic epidemiology of BRCA1 mutations in Norway?

The relevance is that the Norwegian population never was, is not, and never will be, even close to Hardy-Weinberg equilibrium. The relevance is that it is conceptually wrong to address the problem and make a description under the paradigm of balance. The relevance is to dismiss balance as the paradigm and make the description based on a better understanding of the driving forces of the evolution – because the genetic epidemiology of BRCA1 mutations in Norway is not created by survival of the fittest, it not an effect of Darwinism, and we are not in a steady state where nothing will happen if we preserve the environment as is. We are in a state of imbalance, and there are major driving forces which are about to alter the distributions as a consequence of the imbalance. We are in the middle of an evolution not based on selection. Applying that as a paradigm, the current situation is easily understood and some prediction can be made for the future. The basic prediction is that there will be changes because we are not in the state of balance, but what the changes will be may be a stochastic chance and not an effect of selection. From this follows the second paradigm: There is no understanding of why the situation is as is, it is made out of imbalance and is the effect of a stochastic chance, not a selection of the fittest. The question why today's distribution and why not another, has no meaning.

Population geneticists have baptised the concept 'genetic drift'. It is commonly considered a confounder to balance – it may rather be the other way around. Genetic drift is the rule, selection of the fittest is slightly and slowly modifying the results.

Back to Norway – what are the major determinants of today's genetic epidemiology of any rare genetic marker in our population? They are well known: 10,000 years (400 generations) ago there was but ice. A small number of immigrants from central Europe populated the land and made an inbred population in a genetic isolate. This isolate arrived at about 500,000 persons some 1000 (40 generations) years ago, at which time the weather was warm and the naked mountains of today were large forests. At the time we organized the society to the Vikings who military ruled Northern Europe. Coincidencing in time with three major events – new social structure including Christianity, detoriation in weather conditions, and the bubonic plagues – the population collapsed to about 130,000 25 generations ago. The next 15 generations the remaining population lived in small genetic isolates, not broken until modern communications arrived some 5 generations ago. During the 15 generations, the population expanded by inmarriage to several millions, and half the population at the western coast emigrated to USA and Canada to avoid death by starvation about 5 generations ago. Counting the successors of these emigrants, 130,000 have become 10 million without intermarriages between the small isolates until 5 generations ago.

Where does this leave us? It leaves us with the understanding that any dominantly inherited disorder to be relatively frequent in the population, is to be caused of a high mutation rate and/or genetic drift caused by the most recent population bottleneck. It also leaves us with the understanding that – because it is about drift and not selection – there is no way of under-

standing why the present result and not another distribution, because it is the result of chance (the possibility of the Bubonic plagues to select for HLA phenotypes is not discussed here).

Because BRCA1 infrequently mutates, it means that any high prevalence of inherited breast ovarian cancer in Norway is to be caused by a few survivors of the plague having numorous offspings. At the time BRCA1 and BRCA2 were described, we knew that inherited breast ovarian cancer was frequent and unevenly distributed in Norway. Because it was soon clear that the genes rarely mutatate, we actually knew that the high and uneven distribution of breast-ovarian cancer kindreds was the function of genetic drift after the plagues. The task was actually trivial: to demonstrate which mutations, their origin and their present distribution.

How many mutations were expected to survive the plagues? Yates calculated that in a stable population, less than 10% of new mutations will survive 15 generations. In a population of 130,000 and given a population prevalence of BRCA1 mutation carriers of 2/1,000, no more than 260 mutation carriers were assumed to survive the plagues. Because of the mathematical (not selection) loss of infrequent mutations, no more than about 20 of these were assumed to be present today. Obviously, the surviving number may be higher because of the expanding population, but the calculations tell us that we are actually looking for a few frequent mutations, and close to all persons with the same mutations are assumed to be descendants of one mutation carrier surviving the plagues. This because no mutation introduced into the population the last 10 generations will have had any chance to become frequent, the determinants of the present genetic epidemiology were who died during the plagues, and what happened the first generations afterwards.

Reasoning along these lines was in conflict with the current paradigm of Hardy-Weinberg equilibrium, and was not well received in the scientific society. I had to find a lab willing to demonstrate the frequent mutations as a project, and the lab was found in Groningen, The Netherlands because their statistician and myself shared an interest in haplotyping.

To find new mutations in the BRCA genes is expensive, and was even more so at the time. We did not have to, however, because the reasoning told us that all persons with the same mutations were in the same family. So we used VNTRs and haplotyped families from the high incidence areas. It was soon evident that all families had intrafamilial haplotype sharing, and most family at the South west coast shared one out of three haplotypes, while a forth was common in the eastern region. Then we selected four obligate carriers within four families with one of the four haplotypes, and – because all four mutations recided within BRCA1 exon 11 - demonstrated the four dominating founder haplotypes by sequencing exon 11 in four persons only.

We demonstrated that all persons with the same mutation in Norway, share the complete 5cM haplotype within where BRCA1 is located. That is, we have demonstrated the prescribed way by making a hypothesis based on the paradigm to be right. Which is considered epidemiological evidence. We conclude that the current distribution observed for BRCA1 mutations in Norway is caused by the instability produced by the Bubonic plagues 25 generations ago.

We now have a detailed map of the distribution of the locally arisen founder mutations. The four first described mutations account for more than half of all Norwegian mutation carriers. We have developed rapid high capacity test to demonstrate the most frequent mutations (multiplex PCR directly showing the mutations by short DNA segment variations, because 9 of the 10 most frequent mutations are small insertions or deletions). That is, we have a quick-test to

demonstrate about 75% of the mutation carriers in the population. Whether or not to do so, is a political social decision, we have described the epidemiology and constructed the tool to identify the mutation carriers within short and with low cost.

Among the unpleasant experiences along the road, was the negative comments when approaching the paradigms to solve the problems, that no lab was really interested in joining us in the start, which lead us to the situation that it took a year from the results were actually ready until the data was clean enough to publish. In the meantime, others demonstrated by mutation analyses some Jews to have some frequent occurring mutations, which by now are known to have been married into Jews staying in East Europe for a while. Actually, all the old populations in Europe have now demonstrated private founder mutations – possibly besides East Anglia which was never really broken down by the Bubonic plagues. All frequent mutations in local populations have, to the extent they have been examined, been demonstrated to reside on similar haplotypes conform to the paradigm that they all have been made common by genetic drift. By and large, there are but private mutations in BRCA1 world-wide, all frequent mutations represents large, old kindreds.

Today, the story continues by SNPing, because there are more to inherited breast cancer than BRCA1 and BRCA2. The basic paradigm to construct fruitful hypotheses to describe the epidemiology continues to be the opposite of stability.

In general, the world is made by major events which are unforeseeable but we know they have occurred and we know they will occur. These events creates deep instability, and the current situation is but snapshot of an imbalanced system in continuously evolution because of the dynamics in the instability. The current paradigms of preserving a nature in balance, are profound misconceptions. Approaching these paradigms today evokes the same emotional negative energy as they met only a few years ago, when they were introduced.

Underlying is the lack of reflection that any paradigm and any interpretation of scientific results, are but imperfect frames of explaining what we so far know. There is no truth, because we do know that it will change tomorrow when new paradigms will be found more fruitful to describe tomorrow's knowledge. Then the true paradigms of today will be filed as ancient imperfections or misunderstandings. We are as scientists following the path described by Descartes: 'Any problem should be made as simple as possible', with the addition of Einstein: ' – but not simpler'. On a different level, we are continuously exploring the content of Elliot's remark:' Where is knowledge lost in information, where is understanding lost in knowledge, where is wisdom lost in understanding.'

The verbal contribution will give some details from the scientific reasoning and data, arriving at today's concepts when applying SNPs to explore different levels of associations under different paradigms of expected findings.

None of the arguments discussed about are new, and one of the interesting question, therefore, is why the reasoning is to some extent in conflict with current paradigms.

References to the paradigms and data from our studies are given in :

Møller P, Heimdal K, Apold J, Fredriksen Å, Borg Å, Hovig E, Hagen A, Hagen B, Pedersen JC, Mæhle L, The Norwegian Inherited Breast Cancer Group, The Norwegian Inherited Ovarian Cancer Group. Genetic epidemiology of BRCA1 mutations in Norway. Eur J Cancer 2001; 37: 2448-2434.

Gene expression in cancer: clinical impac**t**

Anne-Lise Børresen-Dale
Department of Genetics, Inst. for Cancer Res.,
The Norwegian Radium Hospital, Oslo, Norway

Microarray technologies, applied to the study of DNA, RNA, can be used to portray a tumor's detailed phenotype in its unique context, and to generate molecular signatures that will improve our understanding of the causes and progression of the disease, for the discovery of new molecular markers, for therapeutic intervention and for developing new prevention strategies. Eventually, advances in tumor portraiture will naturally lead to improved and individualized treatments for cancer patients.

We have performed expression studies and genome wide copy number analyses of more than 300 breast carcinomas and 100 ovarian carcinomas aiming at novel tumour classification that can predict survival and treatment response. The expression patterns observed provided a remarkably distinctive molecular portrait of each tumour.

Breast tumours are very heterogonous, but by expression profiling they could be classified into five novel subtypes (two luminal epithelial derived ER+ tumour subtypes, a basal epithelial-like, an ERBB2+ group, and a normal breast-like group). Survival analyses showed significantly different outcome for patients belonging to the various subtypes, including a poor prognosis for the basal-like and a significant difference in outcome for the two luminal /ER+ subtypes. Differences in *TP53* mutation frequency between the subtypes indicated an important role for this gene in determining the gene expression pattern in the various tumors. The frequency of the different genotypes of the codon 72 polymorphism of the *TP53* genes was significantly different in the 5 subgroups identified by expression analyse, indicating that the TP53 genotypes also have an impact on the expression profile. Unequal distribution of the different genotypes of *CYP19* gene between the different subgroups was also observed. Cluster analyses of two published, independent data sets representing different patient cohorts from different laboratories, uncovered some of the same breast cancer subtypes. In the one data set that included information on time to development of distant metastasis, subtypes were associated with significant differences in this clinical feature. By including a group of tumours from *BRCA1* carriers in the analysis we found that this genotype predisposes to the basal tumour subtype. Invasive lobular carcinoma (ILC) is the second most common type of breast cancer next to invasive ductal carcinoma (IDC). We conducted gene expression studies of 21 ILCs and 38 IDCs and identified a subgroup of ILCs ("typical" ILCs) that clustered together and displayed different gene expression profiles from IDCs, while the other ILCs ("ductal-like" ILCs) were distributed between different IDC subtypes. Many of the differentially expressed genes between ILCs and IDCs code for proteins involved in cell adhesion/motility, lipid/fatty acid transport and metabolism, immune/defense response, and electron transport. Many genes that distinguish typical and ductal-like ILCs are involved in regulation of cell growth and immune response. Our data strongly suggest that over half the ILCs (called "typical" ILCs) differ from IDCs not only in histological and clinical features, but also in global transcription programs. The remaining ILCs closely resemble IDCs in their transcription patterns. Different molecular subtypes of ILC may require different therapeutic strategies.

Ovarian cancer is the leading cause of gynecological cancer-related deaths. DNA copy

number alterations (CNAs) are a consistent feature of ovarian carcinoma. cDNA microarrays with 41, 805 features (42K) representing 27, 286 unique genes were used to study CNAs. The results were analyzed using two programs, CGH-Miner and CGH-Explorer. Intervals of CNAs (ICNAs) were identified using Fix-Order clustering, a modified agglomerative clustering method. ICNAs were found at several sites previously reported in ovarian cancer, and the frequency ranged from 0 to 56% with the most frequent gain residing at 8q11.1-8q24.3 (40%) and the most frequent loss at 4q11-4q35.2 (56%). Two novel ICNAs, gain of 19p13.3-19p12 (27%) and loss of 15q13.3-15q22.31 (29%) were identified. Based on variations in patterns of CNA, patients were clustered into three groups, with significant difference in survival. Six amplified ICNAs, including ICNA at 17q12-17q23.2, and a cluster of co-amplified genes located at 17q12, including the oncogene *ERBB2*, were significantly related to poorer disease-specific survival.

These findings set the stage for future studies aimed at identifying genotypes influencing the specific patterns of gene activations that predict important clinical features, like sensitivity to specific therapies and metastatic potential.

References:

Perou, C. M et al. Nature 406: 747-52, 2000
Sørlie, T. et al PNAS 98 (19): 10869-74, 2001
Sørlie, T. et al. PNAS., 100 (14): 8418-8423, 2003
Zhao H. et al. Mol. Biol. Cell 15(6); 2523-2536, 2004
Langerød et al 2005 PloS Medicine, to be resubmitted after revision
Wang et al 2005, submitted

Computational identification of disease related non-coding RNAs.

Ravi Sachidanandam,
Cold Spring Harbor Laboratory, USA

RNAs that are not translated into proteins are called non-coding RNAs (ncRNA) [1,2]. ncRNAs are important components of regulatory networks and are known to be involved in a variety of processes, such as splicing, X-inactivation and RNAi. ncRNAs offer sequence specificity through hybridization. The role of ncRNAs in disease is of great interest as it might point to mechanisms hitherto not considered.

The recently discovered class of ncRNAs, which is of great interest due to its involvement in the RNAi pathway, is the microRNA (miRNA)[3]. There are approximately 250 miRNAs in the human genome [4]. microRNAs are known to silence genes post transcriptionally (PTGS) by either blocking translation or by cleaving the message. The RNAi pathway has also been implicated in transcriptional gene silencing (TGS) through RNA-hairpin directed heterochromatin formation. Thus, compared to transcription factors and other protein based gene control mechanisms, microRNAs and other ncRNAs offer greater sequence specificity and variety in control mechanisms.

Many genetic diseases or predisposition to them, are the result of mutations, single nucleotide polymorphisms (SNP) or insertions/deletions (indels). These indels and SNPs can either affect ncRNA promoters, or the genes themselves, or their binding sites. The basic problem to be solved is one of mapping SNPs/indels to the genome and in parallel:

1) mapping mRNA genes to the genome.

2) mapping ncRNA binding sites on mRNAs and the genome.

There are several studies that have identified SNPs/indels associated with certain cancers [5]. In addition, sequencing of genes from tumors has identified sporadic mutations that seem to have an effect on the response of tumors to drugs. The work we discuss below allows exploration of the effect of these SNPs and indels on genetic pathways through their effect on ncRNAs.

We address the following problems in this talk

1) The management of genomic information. Unfortunately, it is a nightmare to now access genomic data in a biologist friendly way, through names that are familiar to biologists. We solved this problem by developing a tool called GeneSeer which is now available to the public through a website (http://geneseer.cshl.org).

2) miRNA binding sites and cleavage sites. We have developed methods of identifying miRNA binding sites, based on experimental data and identified miRNA binding sites that could result in cleavage of mRNAs.

3) Mapping SNPs and indels to miRNA genes and their binding sites and quantifying their perturbative effects.

4) Mapping ncRNA binding sites to regions of the genome, which could in turn predict effects on heterochromatin formation.

5) Understanding pathway control mechanisms using data from diverse sources, such as homologues across species and miRNA expression microarray data. This in turn can help identify the effect of SNPs and indels on pathways.

References

1. Sean Eddy, *Non-coding RNA Genes and the Modern RNA World.* http://www.euchromatin.org/Eddy01.htm

2. **Database of non-coding RNA** http://noncode.bioinfo.org.cn/index2.htm

3. He L, Hannon GJ. *MicroRNAs: small RNAs with a big role in gene regulation.* **Nat Rev Genet.** 2004 Jul;5(7):522-31.

4. **The miRNA registry**
http://www.sanger.ac.uk/Software/Rfam/mirna/index.shtml

5. Mitra N, Ye TZ, Smith A, Chuai S, Kirchhoff T, Peterlongo P, Nafa K, Phillips MS, Offit K, Ellis NA. *Localization of cancer susceptibility genes by genome-wide single-nucleotide polymorphism linkage-disequilibrium mapping.* **Cancer Res.** 2004 Nov 1;64(21):8116-25.

Dissection of molecular pathways of cancer by high-throughput biochip technologies and RNA interference.

Olli Kallioniemi,
Medical Biotechnology, VTT Technical Research Centre of Finland,
and University of Turku, 20521 Turku, Finland

Our aim is to identify new molecular targets and mechanisms for therapeutic intervention in cancer. To achieve this aim, we develop and apply multiple high-throughput technologies including "in silico" screening as well as technologies for molecular, cellular and clinical discovery research. Finally, data integration from these technology platforms is applied to facilitate interpretation and prioritization of the findings.

In silico screening: In order to make use of the exponential increase of published data on gene expression arrays, we have launched a project to acquire and make use of these data as a discovery resource. We currently have data on 5700 samples analyzed on the Affymetrix gene expression platform stored in our relational database. These samples include e.g. 64 normal tissues/cell types, 43 tumor types, many other diseases as well as functional experiments, altogether 84 Million datapoints. We have developed methods to mine these data to identify tissue- and disease-specific expression patterns of transcripts, gene coexpression profiles, to explore networks of gene regulation as well as methods to interpret new microarray experiments. In silico transcriptomic screening makes it possible to generate dozens of testable hypotheses for laboratory analysis based on data sets that are much larger and more extensive than any single academic laboratory can afford to generate. Analysis of gene expression profiles across 100s of tissue and tumor types, diseases and experimental manipulations generates novel, often unexpected insights on gene function as well as on the underlying biology and medicine.

Molecular screening: Large cohorts of clinical samples are now being investigated not only at the RNA level by gene expression profiling, but also at the DNA-level using comparative genomic hybridization (CGH) arrays for analysis of somatic genetic alterations or SNP-arrays for studies of allelic gains and losses. There is also an emerging interest for large-scale proteomic and metabolic profiling. It will be increasingly important to integrate multiple levels of molecular profiling data to gain new insights and comprehensive views on mechanisms of cancer development. We are applying single-gene resolution oligo-CGH arrays and integrate these data with gene expression information on the same samples. The increased CGH resolution has highlighted several microdeletions as well as small amplifications, whose impact on gene expression can be substantial and highly specific. This has led to an opportunity for rapid identification of genes that may be targets of genetic alterations in cancer. As demonstrated by several recently approved drugs for cancer, such mutated genes represent attractive targets for the development of effective cancer-specific therapeutics.

Functional screening using RNA interference: The molecular profiling of DNA-, RNA- or protein expression patterns in patient samples is not sufficient for implicating these molecules or molecular mechanism as therapeutic targets. It is also necessary to generate functional information on such genes and pathways. Towards this aim, we have developed a high-throughput screening (HTS) system which is composed of a robotic, automated platform for the

analysis of up to 20,000 functional experiments with living cells at a time using the 384-well microplate format. Cells are dispensed into culture wells, exposed to siRNAs or small molecule compounds, incubated for 1-3 days, washed, and stained with phenotype-specific markers for cell growth, cell cycle distribution or induction of apoptosis. The results are read by plate readers or cell cytometers. Functional studies with RNAi libraries (e.g. 1000-10,000 siRNAs) have implicated genes whose targeting by RNAi is lethal to specific cancer types, such as breast cancer. Integration of such functional RNAi data with gene expression and aCGH data has enabled us to identify genes that are targets of genetic alterations and whose expression is required for the maintenance of the malignant phenotype. Such genes represent attractive candidate drug targets.

Clinical screening: Data on molecular targets arising from functional in vitro studies need to be corroborated in studies of large-scale clinical sample cohorts in order to verify that such molecular targets are relevant in clinical patient samples. A number of technologies are being developed towards this aim. First, the in silico screening transcriptomics database with 5700 samples has made it possible to develop an approach for "in silico clinical validation". It is possible to determine the expression levels of any gene across a very large number of tumor and normal sample types. Second, more established technologies, such as tissue microarrays (TMA) facilitate the analysis of individual DNA, RNA and protein targets in thousands of arrayed patient samples, typically from formalin-fixed tumors. TMA analysis with antibody binders produces definitive clinical data on the expression of therapeutic protein targets, and enables quantitation of drug target distributions at the population level (target epidemiology). In order to further increase the throughput of molecular analyses, we are developing methods to print tissue lysates from frozen samples in an array format. Tissue protein lysate arrays will help to automate and expand antibody-based validation of molecular targets in large cohorts of tissue specimens.

In summary, high-throughput technologies can be applied in the "in silico", molecular, cellular and clinical screening studies, thereby substantially increasing the information and insights that can be obtained on a single platform. Integration of data from the various high-throughput platforms will facilitate a deeper, mechanistic interpretation and understanding of cancer biology, which is an essential requirement for the development of next-generation targeted therapies.

Gene expression and radiation sensitivity

OK Rødningen[1], J Alsner[2], T Hastie[3], J Overgaard[2] and A-L Børresen-Dale[1]

[1]Department of Genetics, The Norwegian Radium Hospital, Norway, [2]Department of Experimental Clinical Oncology, Aarhus University Hospital, Denmark, and [3]Department of Statistics, Stanford University, USA

Among breast cancer patients receiving ionizing radiation (IR) treatment, a subgroup show long-term adverse effects in the normal tissue. Women treated for unilateral breast cancer may also develop breast cancer in the contralateral breast, possibly caused by the radiation exposure. Despite that a linear relationship between radiation dose and IR response have been documented, a large variation in the adverse effect to the treatment is observed among patients receiving the same dose. This enhanced sensitivity to the damaging effects of IR among a subset of patients support the hypothesis that the variation in response is genetically determined.

One of the most common adverse effects is radiation-induced fibrosis (RIF), a complex biological process developing gradually over several years (O'Sullivan and Levin, 2003). RIF is believed to occur as the result of a coordinated response to IR involving several different cytokines and growth factors, fibroblast proliferation and differentiation, and also remodelling of the extracellular matrix . Risk of RIF is at present a dose-limiting factor in radiation therapy. The mechanisms whereby the initial cellular and extracellular responses following irradiation of the skin promotes RIF are not fully understood. However, it is known that IR induces DNA double-strand breaks (DSBs). These DSBs activate a serine/threonine protein kinase; ATM, either through changes in chromatin structure (Bakkenist and Kastan, 2003) or through Mre11-Rad50-Nbs1 mediated unwinding of DNA ends (Lee and Paull, 2005). ATM phosphorylates several downstream molecules including p53, one of the key proteins in the checkpoint pathways. Its activation leads to the transcription of several genes of which the products can trigger processes such as cell cycle arrest, apoptosis, DNA repair and differentiation. In addition, the generation of reactive oxygen species by IR can cause damage to both DNA, proteins and membrane lipids, initiating further stress responses in the cell. Thus, IR initiates a cascade of events including induction of several signal transduction pathways, and major shifts in the transcription profiles of cells.

The fibroblast is one of the key players in the process of radiation-induced subcutaneous fibrosis. To increase our understanding of the effects of IR on human tissue and cells, we have used cDNA microarrays (www.mikromatrise.no) to analyze the transcriptional responses of in vitro radiated fibroblasts. Thirty different human fibroblast cell lines were included in the study. The fibroblasts were achieved from patients treated with post-mastectomy radiotherapy in Aarhus, Denmark, from 1978-1982, and subsequently evaluated in detail with regard to development of RIF (Johansen et al., 1996). Cells were irradiated with either single doses of 3.5 Gy or multiple doses of 2 Gy. Expression analyses were performed on unexposed and exposed cells after different time points. The IR response was analyzed using the statistical method Significance Analysis of Microarrays (SAM, www-stat.stanford.edu/~tibs). We identified known

IR responsive pathways, in which several seem to be mediated by ATM and p53. We also identified responses not previously described, which might be associated with IR-induced fibrosis. SAM analyses were also applied to identify genes in which the expression level correlated with the level of fibrosis, and PAM analyses (Prediction Analysis for Microarrays, www-stat.stanford.edu/~tibs) identified a limited set of predictive genes. These genes may provide basis for a diagnostic tool in the identification of patients with adverse responses to radiation, and improve and optimize radiotherapy at the individual level.

References:

Bakkenist, C.J. and Kastan, M.B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. Nature *421*, 499-506.

Johansen, J., Bentzen, S.M., Overgaard, J., and Overgaard, M. (1996). Relationship between the in vitro radiosensitivity of skin fibroblasts and the expression of subcutaneous fibrosis, telangiectasia, and skin erythema after radiotherapy. Radiother. Oncol. *40*, 101-109.

Lee, J.H. and Paull, T.T. (2005). ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. Science *22*; 308(5721), 551-554.

O'Sullivan, B. and Levin, W. (2003). Late radiation-related fibrosis: pathogenesis, manifestations, and current management. Semin. Radiat. Oncol. *13*, 274-289.

Supported by EMBIO, University of Oslo, Norway and the Danish Cancer Society.

A Global Overview of Human Genetic Variation

Kenneth K. Kidd, PhD, Yale University, USA

One of the grand challenges of the post−human-genome-sequence era is to "develop a detailed understanding of the heritable variation in the human genome". By characterizing genetic variation among individuals and populations, we may gain a better understanding of differential susceptibility to disease, differential response to pharmacological agents and the complex interaction of genetic and environmental factors in producing phenotypes. A fundamental truth is that allele frequencies vary among population and that the variation generally follows a geographic pattern. This roughly corresponds to the lay person's concepts, because people from different parts of the world look different, whereas people from the same part of the world tend to look similar. However, simplification to "five races" ignores both the complexity of the global pattern as well as the variation within each geographic region. Both will be illustrated in this talk.

The genetic data accumulated over the past two decades overwhelmingly support the Recent African Origin (RAO) model (also called the Out of Africa model). According to the RAO model, all non-African populations descend from an anatomically modern *H. sapiens* ancestor that evolved in Africa ~200 thousand years ago (Kya) and then spread and diversified throughout the rest of the world starting ~50−100 Kya, supplanting any archaic *Homo* populations still present outside of Africa, such as Neanderthals (although low levels of admixture between these groups cannot be ruled out). Studies of variation in autosomal, mitochondrial DNA (mtDNA) and Y-chromosome haplotypes indicate that the migration(s) out of Africa originated from an East African gene pool. The RAO model predicts a recent common African ancestor with subsequent recent expansions after the initial migration(s) out of Africa ~100 Kya. This history of expansions out of Africa and then into Australo-Melanesia (~60 Kya), Europe (~40 Kya), Asia (~35 Kya), the New World (~30 Kya) and the Pacific (~3 Kya) is a major cause of the global patterns of allele frequency variation.

Amounts of genetic variation Humans are ~98.8% similar to chimpanzees at the nucleotide level and are considerably more similar to each other, differing on average at only 1 of every 500−1,000 nucleotides between chromosomes. This degree of diversity is less than what typically exists among chimpanzees. Current estimates of how much variation occurs species-wide indicates that all *H. sapiens* are ~99.6−99.8% identical at the nucleotide sequence level. The other 0.2−0.4% of 3 billion nucleotides comprises ~10 million DNA variants that can potentially occur in all different combinations (these numbers may be underestimates, as Build 124 of dbSNP already contains nearly 10 million single-nucleotide polymorphisms (SNPs), of which 5 million are validated;

Global patterns of genetic variation

Studies of autosomal and X-chromosomal haplotype variation, as well as mtDNA variation, indicate that Africans have the largest number of population-specific alleles and that non-African populations carry only a fraction of the genetic diversity that is present in Africa. This would be

expected if there were a genetic bottleneck at the time of migration of modern humans out of Africa. We also see a clinal pattern of less heterozygosity and fewer population-specific alleles with increasing distance from Africa. This would be expected from accumulating random genetic drift at the forefront of expansion with "stabilization" of the frequencies as populations increased in size behind the expansion front. Populations also differ with respect to the organization of variants along a chromosome (haplotypes). The nonrandom association of alleles at different sites is referred to as linkage disequilibrium (LD). Levels and patterns of LD depend on gene-specific factors, such as selection and rates of mutation and recombination, as well as demographic factors that have a genome-wide effect, such as population size, population structure, founder effect and admixture. At large distances (multi-mega bases) recombination is sufficiently frequent that LD is not seen in any population. At shorter distances recombination becomes a stochastic factor that in combination with random genetic drift determines whether or not alleles at nearby sites show LD. Numerous studies of LD between SNPs and microsatellites show greater LD in Eurasians than in Africans and still greater LD in Native Americans. Additionally, levels of LD may vary within geographic regions. This pattern of LD is consistent with human demographic history; ancestral African populations have maintained a larger effective population size (N_e) and have had more time for recombination and mutation to reduce LD. The bottleneck associated with the expansion of modern humans out of Africa resulted in many of the African haplotypes being lost, leading to greater LD in non-African populations. Another bottleneck, associated with the expansion into the Americas, is reflected in the even higher amounts of LD in this region. Outside of Africa the time has been too short for recombination to randomize alleles at nearby sites. And, because of drift, one cannot readily estimate recombination rate from the frequency of a recombinant haplotype.

A current research question is whether the genome is organized in blocks. Our data argue that the blocks seen in European populations are not fundamental characteristics of the human genome but statistical artifacts of haplotype frequencies. As noted above, the positions at which LD breaks down because of a high frequency of recombinant haplotypes may represent only a single rare crossover product, analogous to a mutation, that has drifted to a high frequency in subsequent generations. Even among European populations the patterns of "blocks" can differ because of different frequencies of the same few haplotypes. On a global basis the effect is even greater. Thus, the block concept is not generally useful and blocks defined in one population will not necessarily apply in another population. However, patterns of LD can be used to identify the minimum set of markers (usually single nucleotide polymorphisms, SNPs) that identify and distinguish the major haplotypes in a population. These tagging SNPs can save considerable laboratory effort while preserving the information on the haplotype lineages in the population. Our data show, however, that with differences in haplotype frequencies the optimum set of tagging SNPs can differ. Thus, definition of tagging SNPs is population specific and sets defined in one population will not necessarily generalize to other populations even from the same region of the world.

From families to populations: impact of disease genes

Leena Peltonen, M.D., Ph.D. Professor,
Department of Molecular Medicine, National Public Health Institute and Department of Medical Genetics, University of Helsinki, Finland; Department of Human Genetics, David Geffen School of Medicine at UCLA, USA

The Human Genome Project has produced a high number of catalogued sequence variants enabling genome-wide studies of genetic loci behind rare- and common disease-related phenotypes. However, the strategy of disease gene hunt is highly dependent on the accessible population and study samples. Multiple uncertainties must be solved before the best possible strategy for the gene hunt can be designed and large-scale genome-wide investigations undertaken. Further, rapidly increasing information of the structural or functional variability within the genome will greatly affect the interpretation of the impact of identified variants.

European population isolates like Finland have been very useful for mapping and cloning genes for rare disorders; in such isolates genetic drift leads to an overabundance of disease-alleles for particular disorders, and a high proportion of patients share these alleles, identical by descent. The concept that the isolates are similarly advantageous for genetic studies of common diseases has been challenged, and only few samples exist to prove if it really would be more straightforward in such populations to detect disease-related haplotype signatures through association studies. Despite these reservations, detailed information of the population history is increasingly understood as one crucial factor of success in genetic studies of common diseases. I will describe the features of the Finnish population and our efforts to search for disease genes for rare and common phenotypes. Our studies will exemplify the strategies used to identify disease genes in various Finnish data sets and study samples. I will also demonstrate how the impact of specific allelic variants of disease genes, identified in families, can be addressed in epidemiological cohorts containing excessive amount of quantitive phenotype information. I will also describe our recent efforts in the EU-funded multinational twin study, Genomeutwin (www.genomeutwin.org).

Identification of Low Penetrance Genes and Their Impact on Disease Risk

David R. Cox M.D., Ph.D.
Chief Scientific Officer
Perlegen Sciences Inc.
Mountain View, CA

We describe the first detailed analyses of the structure of common human genetic variation on a whole genome scale. The primary data that for these analyses are genotypes of 1.6 million SNPs in 71 individuals from three populations with substantially different histories. We find that a large fraction of all common human variation is captured by a small fraction of all common SNPs. Since the samples investigated in this study do not capture the full genetic diversity of the populations from which they were selected, our data is not suitable for answering many questions about the detailed genetic structure of human populations. However, the public availability of these data will enable a wide variety of additional analyses to be carried out by scientists investigating the structure of human genetic variation as well as the genetic basis of human phenotypic differences.

Our focus on common genetic variation has several motivations. Common variants account for a larger share of human nucleotide diversity than rare variants and are more experimentally tractable. For the same allelic effect, a common variant represents a larger fraction of phenotypic variance and population attributable risk than a rare one, so common variants are more valuable from the perspective of diagnostics and intervention. Finally, detecting and characterizing effects of rare variants requires very large sample sizes to obtain statistically meaningful numbers of individuals carrying a rare allele. There is no doubt that rare variants play a role in the etiology of common disease, but pursuit of common variants is more tractable with available technologies.

Common human diseases, such as cardiovascular disease and psychiatric illness, are caused by the interplay of multiple genetic and environmental factors. The bounded nature of the human genome and the availability of the complete human genome sequence have resulted in extensive efforts to define the genetic basis of a wide variety of complex human traits. One approach for identifying such genetic risk factors is the case-control association study, in which a group of individuals with disease is found to have an increased frequency of a particular genetic variant compared to a group of control individuals. A number of genetic risk factors for common disease have been identified by such association studies. These studies suggest that many different genes distributed throughout the human genome contribute to the total genetic variability of a particular complex trait, with any single gene accounting for no more than a few percent of the overall variability of the trait. Case-control study designs employing on the order of 1000 total individuals can provide adequate power to identify genes accounting for only a few percent of the overall genetic variability of a complex trait, even using the very stringent significance levels required when testing large numbers of common DNA variants. Using such study designs in conjunction with a detailed description of common human DNA variation, it may be possible to identify a set of major genetic risk factors contributing to the variability in a complex disease and/or treatment response. While knowledge of a single genetic risk factor can seldom be used to predict treatment outcome of a common disease, knowledge of a large fraction of all the major genetic risk factors contributing to a treatment response or common disease could have immediate utility, allowing existing treatment options to be matched to individual patients without requiring additional knowledge of the mechanisms by which the genetic differences lead to different outcomes.

Genetics of migraine

Aarno Palotie M.D., Ph.D.
Finnish Genome Center, University of Helsinki, Finland

Migraine is a common, complex neurovascular disorder. Migraine has two main types: migraine without aura (MO) occurring in 85% of patients and migraine with aura (MA) occurring in 15% of patients. Despite the old notation that migraine runs in families little is known about its molecular genetic background. Family and twin studies have subsequently indicate a genetic component and a multifactorial inheritance. The only true molecular insight in migraine pathophysiology is provided by the identification of two genes underlying rare, Mendelian forms of migraine, Familial Hemiplegic Migraine (FHM). Both of these genes are involved in the ion transport of the cell; the calcium channel gene (CACNA1A) and the alpha subunit of the Na/K pump gene (Na/K-ATPase). There is no conclusive evidence that variants of these two genes would contribute to the susceptibility of more common forms of migraine. However, CACNA1A knock in mouse experiments demonstrate similar cortical spreading depression as recorded during migraine attacks in human. Genome wide linkage studies of common forms of migraine have revealed linkage to a few chromosomal regions, but no associated gene variant has been reported as yet. Our strategy has been to collect most likely the largest migraine family cohort, which we are using for gene identification by a positional cloning strategy. It includes samples and detailed phenotype information from almost 700 families and more than 4500 individuals. These families have been used to identify susceptibility loci using genome wide scans and subsequently to restrict linked regions. The large family sample allows also to construct case control studies with 1000 affect MA cases and carefully matched controls. To provide an alternative strategy to identify positional and functional candidate genes, we have carried out genome wide expression studies from CACNA1A mutant mice and their wild type litter mates. These candidate genes can then be correlated with linkage and association results from human family and case control studies. Subsequently identified loci will be tested in large epidemiological cohorts including the twin project GenomEUtwin consisting of over 800 000 twin pairs in eight countries as well as the EuroHead project, a collaboration of eight European headache centers with multiple headache cohorts.

THE MOLECULAR BASIS FOR OAT INTOLERANCE IN PATIENTS WITH CELIAC DISEASE

Knut E. A. Lundin for Celiac Research Group (leader professor Ludvig M. Sollid), Institute of Immunology and Department of Medicine, Rikshospitalet, Oslo, Norway

Celiac disease (CD)

CD affects 1:200 to 1:100 in all Caucasian populations. It may occur at any age after introduction of gluten in the diet. It is characterized by small intestinal inflammation resulting in crypt hyperplasia and villous atrophy, and epithelial and mucosal accumulation of activated T and B cells. The clinical manifestations vary considerably; iron deficiency, osteoporosis and fatigue are typical features. Malignancies occur, both lymphomas and adenocarcinomas.

Genetic background

CD is strongly associated to particular HLA class II variants. More than 95% carry HLA-DQ2 (encoded by DQA1*05/DQB1*02), the rest carry HLA-DQ8 (encoded by *DQA1*03/DQB1*0302*). These HLA alleles are very common also in the general population. There is also a contribution of non-HLA-linked genes, which together may account for a bigger effect than the HLA-linked genes. Genes on chromosome s 2q33, 5q32, and 19p13 are all candidates.

Pathophysiology

Central is the finding of an intestinal T cell response directed against fragments of wheat gluten and related proteins in other cereals. This intestinal T cell response is not present in healthy individuals. The T cells recognize fragments of gluten that are resistant to digestive proteolysis, and which are presented by the disease associated DQ2 and DQ8 molecules. The fragments are typically more immunogenic following deamidation (glutamine to glutamate conversion) by the enzyme tissue transglutaminase in the mucosa.

Wheat, rye, barley and oats

Wheat, rye and barley are closely related and contain similar immunogenic fragments. Oats is the nearest relative; contains less “gluten” and fewer immunogenic fragments. CD is treated by a gluten free diet devoid of these cereals.

Challenge studies

Clinical challenge studies with oats during the last 10 years have enrolled approximately 150 adults and 150 children. Only one of the studies was done double-blinded and controlled. Patient drop-out has been considerable. The conclusion of all but one of the studies was that oats is safe for celiacs.

Intolerant patients

In a clinical challenge study we found one oats intolerant patient. Two were later identified from clinical practice (third-line referral centre).

T-cell immunity to oats

The intolerant patients all had a strong intestinal T cell response towards oats. The T cells were restricted by HLA-DQ2, they recognized a peptide similar to but not identical to those found in wheat. The immunogenicity of the peptide was enhanced by tissue transglutaminase mediated deamidation.

Interpretation

Some celiac disease patients are intolerant also to oats, whereas most are tolerant. This is compatible by different T cell repertoires among groups of patients. This is important information for celiac disease patients who are consuming oats and their clinicians.

ABC Transporters and LXR Receptors: Targets for Cardiovascular Drug Discovery

Richard M. Lawn
CV Therapeutics, Inc.
Palo Alto, CA 94304 USA

Cardiovascular disease is the major cause of mortality in industrialized nations. The low density lipoprotein (LDL) particle in its modified forms [including oxidized LDL and Lp(a)] deposit cholesterol in cells of the vessel wall, including macrophages and smooth muscle cells, contributing to the progression of the disease. Modern drugs are quite effective in reducing LDL levels. However, roughly half of heart disease patients have low levels of high density lipoprotein (HDL), and not elevated LDL. HDL removes excess cholesterol from peripheral tissues via the pathway called reverse cholesterol transport. In order to inform possible drug discovery, researchers sought to discover the defect in Tangier disease, a rare monogenic disorder marked by extreme HDL deficiency and heightened cardiovascular risk.

Utilizing different strategies, several groups identified the defective Tangier disease gene to encode the transporter protein ABCA1. While "traditional" pedigree analysis was used by others to identify the defective locus, we used gene expression microarray profiling followed by cell culture expression studies to establish the role of ABCA1. It was shown to be expressed in macrophages in cardiovascular lesions, and act as a rate limiting step in the efflux of cholesterol and phospholipids onto nascent HDL particles. Genetic evidence in mice and humans verified the sense of increasing ABCA1 gene expression as a drug discovery strategy. The nuclear hormone receptor LXR is a key physiologic modulator of ABCA1 expression, and may offer one of the few possible pathways of pharmacologically increasing ABCA1 activity. Since the initiation of this strategy, numerous LXR target genes have been identified in macrophages, liver, intestine and other sites, which may serve to orchestrate a concerted physiological response to excess sterol deposition. These include at least three other members of the ABC transporter family. Two of them are now identified as agents for another rare genetic disorder of sterol metabolism, sitosterolemia.

Unfortunately, a flaw in this strategy arose with the discovery that systemic administration of potent LXR ligands causes increased plasma triglycerides and liver lipid deposition due to the induction of several gene products involved in the synthesis of fats. Lipogenic genes in the liver are highly induced by LXR activation either directly, or via LXR induced transcription of the sterol regulatory protein SREBP1c. Pharmaceutical development is taking several approaches to obtain LXR agonists that are selective for either tissues, for LXR receptor subtypes, or particular target genes. Macrophage replacement studies in mice have supported the hypothesis that LXR activation confined to macrophages would have a protective role in reverse cholesterol transport but avoid the pitfalls of inducing lipid bio-synthetic genes in the liver. A promising strategy might be to take advantage of the very complexity of the LXR enhancer/promoter transcription complex to discover compounds that would increase transcription of the sterol transport set of LXR target genes, but not the lipid bio-synthetic target genes. This is being approached by both structure-based design of LXR ligands that may perturb the co-activator binding region in different ways, and by a compound screening strategy which incorporates readouts of multiple target genes.

Two good recent reviews on this subject are:

Joseph S.B. and Tontonoz, P. LXRs: new therapeutic targets in atherosclerosis? Current Opinion in Pharmacology 2003, 3: 192-197.
Brewer B.B. et al. Regulation of plasma high density lipoprotein levels by the ABCA1 transporter and the emerging role of high density lipoprotein in the treatment of cardiovascular disease. Arterioscler. Thromb. Vasc. Biol. 2004, 24: 1755-1760.

Title: Apo(a) size heterogeneity and measurement of Lp(a) levels: impact on interpretation of clinical data

Santica M. Marcovina, PhD, ScD
Director, Northwest Lipid Metabolism and Diabetes Research Laboratories, University of Washington, Seattle, WA, USA

In the United States it has been estimated that 37% of people classified to be at high risk for developing coronary artery disease (CAD) based on the guidelines of the National Cholesterol Education Program have increased plasma levels of Lp(a) while only 14% of those judged to be at low risk have high Lp(a) levels. While numerous studies have documented that high plasma Lp(a) concentrations are associated with a variety of cardiovascular disorders, Lp(a) continues to defy the ability of scientists to elucidate its physiologic role and its pathologic mechanisms of action. This, in turn, makes it difficult for clinicians to develop effective interventions in clinical practice. Lp(a) is unique among the lipid risk factors in that it is extremely variable in size and density, its concentrations vary from <0.1 to >200 nmol/L, the risk does not appear to be continuous but increases at high concentrations, and a synergistic interaction of high Lp(a) levels with other lipid risk factors has been reported in several studies. The size and density heterogeneity of Lp(a) are almost entirely due to the genetic size polymorphism of the polypeptide chain of apo(a), the unique protein component of Lp(a). The kringle 4 (K4) type 2 motif is present in a variable number of identically repeated copies giving origin to the 35 apo(a) isoforms detected in human plasma (1). Results of early studies demonstrated that all apo(a) isoforms are co-dominantly expressed at a single locus by multiple autosomal alleles and that the sizes of apo(a) isoforms are in general inversely correlated with plasma Lp(a) concentrations (2-4).

The variable number of identically repeated copies of K4 type 2 of apo(a) has a marked effect on the immunochemical measurement of Lp(a). In fact, based on the number of apo(a) identical repeats, there will be a variable number of epitopes in plasma samples available to interact with the antibodies used in the immunoassays to measure the concentration of Lp(a). The main requisite for an immunoassay to provide accurate results are: (1) assay calibrators should have an accuracy-based target value, and (2) the antibodies should possess the same immunoreactivity per particle for the assay calibrator and for the samples being analyzed. This latter requisite is not easily achievable for Lp(a) and it is evident that the size variation of apo(a) affects the development of immunoassays on two fronts: (1) the choice of apo(a) size in the calibrator is arbitrary and, independent of the choice, the calibrator would not be representative of the size of apo(a) in most of the samples; and (2) the immunoreactivity of the antibodies will vary depending on the size of apo(a). As a consequence, the immunoassays will tend to underestimate Lp(a) levels in samples with apo(a) sizes smaller than the apo(a) size of the assay calibrator and conversely overestimate Lp(a) levels in samples with larger apo(a) isoforms.

A large number of commercially available methods were evaluated using a common reference material to assign a target value to the assay calibrators. The among-method CV on 30 selected samples ranged from 6% to 31% (5) and a significant size-dependent bias was observed for the majority of the assays when the obtained values were compared to those obtained by a reference ELISA method (6). This assay is based on the use of a high affinity monoclonal

antibody that is directed to a unique epitope only present in K4 type 9 of apo(a). By not recognizing the K4 type 2 motifs, this monoclonal antibody accurately measures Lp(a) concentration in plasma irrespective of the different apo(a) isoforms. In the majority of clinical studies, Lp(a) levels have been determined by methods affected by apo(a) size heterogeneity. Therefore, for the conclusions of these studies to be valid, we have to rely on the unlikely assumption that the distribution of apo(a) isoforms is the same between cases and controls, thus minimizing the potential that method-dependent overestimation or underestimation of Lp(a) values may contribute to the observed difference or lack thereof between cases and controls.

We have recently evaluated the impact of method inaccuracy on the correct stratification of individuals for coronary artery disease risk based on their Lp(a) values (7). In 2,940 samples from the participants in the Framingham study collected during the 5th cycle, Lp(a) levels were determined by the reference ELISA method (6) and apo(a) isoforms were determined as previously described (1). During the same cycle, Lp(a) levels were also determined in other laboratories in 2,556 of the samples by a turbidimetric method and in 2,662 of the samples by a commercially available ELISA. An Lp(a) value of 75 nmol/L was selected as the clinical decision point. Based on the Lp(a) values obtained by the turbidimetric assay, 136 individuals were misclassified as being at increased risk (false positive) and 23 were misclassified as being not at risk (false negative). Based on the Lp(a) values obtained by the commercial ELISA, 329 individuals were misclassified as being at increased risk (false positive) and 25 were false negative. The vast majority of the misclassifications observed by the turbidimetric method were explained by the overestimation or underestimation of Lp(a) values based on the apo(a) size in the samples. The larger number of false positives as compared to the false negatives obtained by this assay is explained by the small apo(a) size in the assay calibrator and by the high frequency of samples in the general population with apo(a) sizes larger than that in the calibrator. This observation is consistent with previously reported data (6). However, the large number of false positive values generated by the ELISA method was not explained by the apo(a) sizes in these samples. When this method was evaluated in our laboratory, we observed a high degree of intra- and inter-assay imprecision. Additionally, when analyses were performed on fresh samples and on the same samples stored at -70°C, a significant difference was observed between values obtained on fresh and frozen samples. The values on frozen samples were generally higher than those obtained on fresh samples, but the magnitude of the increase was sample-dependent.

These findings clearly indicate that each assay, in addition to being evaluated for its sensitivity to apo(a) size polymorphism, needs to be optimized in terms of robustness and precision. In addition, assay limitations due to sample preparation and storage need to be evaluated and clearly stated.

To evaluate method differences in the interpretation of clinical data, a study was performed to directly compare the ability of the ELISA reference method and a commercially available latex-based nephelometric method to predict future angina pectoris in men participating in the Physicians' Health Study (8). Plasma Lp(a) concentration was determined in our laboratory by the ELISA reference method and apo(a) isoform size was also determined as previously published (1). In a different laboratory, Lp(a) levels were determined on the same samples by the nephelometric method. Analyses were performed in samples from 195 study participants who subsequently developed angina and from 195 paired matched controls. As determined by the ELISA method, cases had significantly higher median Lp(a) compared to controls (30.6 nmol/L vs. 22.5 nmol/L; p=0.02). Additionally, Lp(a) was associated with increased relative risk for angina and this association was strengthened after controlling for lipid risk factors. No significant difference in median Lp(a) levels between cases and controls was observed when Lp(a) was

measured by the nephelometric method (11.4 mg/dL vs. 8.8 mg/dL; p=0.11). The median Lp(a) level and the association with angina was also not significant. In this study it was also found that small apo(a) size predicts angina with greater strength than Lp(a) concentration, thus suggesting that small apo(a) isoforms, in addition to being associated with higher plasma Lp(a) concentrations, may in and of themselves be more atherogenic.

Large clinical and population studies need to be performed to evaluate the ethnic differences of the relationship between apo(a) size and Lp(a) concentration, to establish guidelines for the use of Lp(a) in clinical practice, and to define its importance in relation to other risk factors for coronary artery disease. To compare data among different studies and populations, it is essential that assays for measuring Lp(a) be validated for their ability to provide accurate values independently of the size of apo(a) in the samples as recommended in the recent report of the National Heart, Lung and Blood Institute (7).

REFERENCES:

1. Marcovina SM, Zhang ZH, Gaur VP, Albers JJ. Identification of 34 apolipoprotein(a) isoforms: differential expression of apolipoprotein(a) alleles between American blacks and whites. Biochem Biophys Res Commun 1993;191:1192-1196.
2. Utermann G, Menzel HJ, Kraft HG, Duba HC, Kemmler HG, Seitz C. Lp(a) glycoprotein phenotypes. Inheritance and relation to Lp(a)-lipoprotein concentrations in plasma. J Clin Invest 1987;80:458-465.
3. Lackner C, Cohen JC, Hobbs HH. Molecular definition of the extreme size polymorphism in apolipoprotein(a). Hum Mol Genet 1993;2:933-940.
4. Utermann G, Duba C, Menzel HJ. Genetics of the quantitative Lp(a) lipoprotein trait. II. Inheritance of Lp(a) glycoprotein phenotypes. Hum Genet 1988;78:47-50.
5. Marcovina SM, Albers JJ, Scanu AM, Kennedy H, Giaculli F, Berg K, Couderc R, Dati F, Rifai N, Sakurabayashi I, Tate JR, Steinmetz A. Use of a Reference Material Proposed by the International Federation of Clinical Chemistry and Laboratory Medicine to Evaluate Analytical Methods for the Determination of Plasma Lipoprotein(a). Clin Chem 2000;46:1956-1967.
6. Marcovina SM, Albers JJ, Gabel B, Koschinsky ML, Gaur VP. Effect of the number of apolipoprotein(a) kringle 4 domains on immunochemical measurements of lipoprotein(a). Clin Chem 1995;41:246-255.
7. Marcovina SM, Koschinsky ML, Albers JJ, Skarlatos S. Report of the National Heart, Lung, and Blood Institute Workshop on Lipoprotein(a) and Cardiovascular Disease: recent advances and future directions. Clin Chem 2003;49:1785-1796.
8. Rifai N, Ma J, Sacks FM, Ridker PM, Hernandez WJ, Stampfer MJ, Marcovina SM. Apolipoprotein(a) size and lipoprotein(a) concentration and future risk of angina pectoris with evidence of severe coronary atherosclerosis in men: The Physicians' Health Study. Clin Chem 2004;50:1364-1371.

Oxidized Phospholipids and Lp(a) Lipoprotein, a role in Innate Immunity?

Claes Bergmark, M.D, PhD.
Department of Vascular Surgery, Karolinska University Hospital and Karolinska Institute, Stockholm, Sweden

Atherosclerosis is characterized as a disease of chronic, low grade inflammation. Inflammation is a major effector mechanism by which immunity eliminates threats. The immune system consists of a detector and an effector part. The detector part can be divided into innate and adaptive immunity, both of importance for the development and clinical events in atherosclerosis. The innate immunity provides a fast and blunt response against highly conserved motifs in pathogens, called PAMPs (pathogen-associated molecular patterns). The most important cell types expressing this response are macrophages and dendritic cells by expressing pattern-recognition receptors (PRRs) as scavenger receptors and toll-like receptors. The ligands for these receptors (PAMPs), consists of a diverse array of compounds, but can often be derived to a part of the bacterial cell wall. Soluble factors of the innate defense are CRP, complement and IgM antibodies produced by B1-cells.

Oxidized phospholipids (OxPL) have wide and important biological effects.(1-4) In minimally oxidized Low Density Lipoprotein (OxLDL), they can form covalent adducts with apoB. These adducts retain the intact phospohorylcholine (PC) headgroup, and are called "neo-self" oxidation specific epitopes.They provide many of minimally OxLDL pro-atherogenic qualities. OxPL also has a role in phagocytosis of apoptotic cells, membrane redox regulation and as a proinflammatory cell signal. In Innate immunity, OxPL-PC is a ligand for scavenger receptors, providing an important mechanism for foam cell formation. The PC-epitope, when presented on OxPL or as a part of the streptococcus pneumoniae cell wall is recognized by the innate immune systems other receptors, as CRP and EO6-IgM.

EO6 is a murine natural monoclonal IgM antibody with specificity for PC presented in specific conditions provided by OxPL or as a component of the capsular polysaccharide bacterial cell wall. EO6 epitopes are present in the atherosclerotic lesion.(5-7)

Lp(a) lipoprotein consists of LDL with an additional protein, called apoprotein a (apo a) attached to LDL by a single disulfide bond to the protein in LDL, apoB100. High plasma levels of Lp(a) is a risk factor for atherosclerosis development and its clinical events, especially when combined with high plasma levels of LDL. The distribution of plasma levels of Lp(a) are highly skewed and genetically determined. This is due to a wide variation of the size of apo a, in turn dependent of number of repetition of a part of this molecule, called kringle IV-2. The physiological function of Lp(a) is not known, but much research addressing this question has apo a's resemblance with plasminogen as starting point.(8)

Measuring plasma levels of OxPL as EO6-positive apoB100-containing molecules in plasma gave the surprising result that there was a very high correlation between Lp(a) and EO6. This was a constant finding in many populations .A first step to reveal the nature of the association between Lp(a) and Ox PL showed that kringle V in apo a could bind 2 molecules of OxPL by lysine residues numbered 12 and 42. Kringle-V has been reported to stimulate Interleukin-8 production in macrophages, an important proinflammatory interleukin. OxPL also stimulate interleukin-8 production, so it could be speculated that the OxPL in kringle V could be contributing to this effect.(9)

In vitro transfer experiments between different classes of apoB100 containing molecules for OxPL showed that there was an enhanced transfer of OxPL to Lp(a), and also that when using Lp(a) as donor, no transfer occurred.(10)

Following patients with plasma samples before and after PCI at different timepoints showed that there was a sharp and immediate rise in OxPL plasma levels after PCI, shortly followed by a rise in Lp(a) levels. Most of the OxPL where found on non-apoB100 containing fractions immediately after PCI, but later most OxPL was found on Lp(a).(11)

The same conditions were found in following patients after an acute coronary syndrome event. There was a rise in both OxPL and Lp(a).(12)

Following patients with statin treatment at baseline and after 16 weeks of treatment (MIRACL trial) showed that total OxPL on all apoB100 particles was decreased, but there was enrichment of OxPL on a smaller pool in parallel with similar increases in Lp(a). This supports the hypothesis that statin treatment promotes mobilization and clearance of OxPL, which may contribute to a reduction in ischemic events.(13)

Following healthy women put on a low total fat diet showed a rise in OxPL and Lp(a) plasma levels. This diet also was rich in antioxidants and polyunsaturated lipids. So, a traditionally "anti-atherogenic" diet showed this unexpected finding, maybe mirroring a similar mechanism proposed for statin treatment with mobilization of OxPL.(14)

There is now a newly developed murine model expressing high levels of Lp(a) (700 mg/dl, compared to the 30mg/dl threshold associated with increased risk for atherosclerosis). A second line of this model has levels of 35 mg/dl. The high level Lp(a) model also has corresponding high levels of OxPL.This model will be of much help in future studies of the relationship and effects between Lp(a) and OxPL.(15)

Returning to the model of Innate Immunity, it could be speculated that Lp(a) acts as a "soluble receptor" for OxPL in the same manner as scavenger receptors and CRP with the purpose to "detoxicate" tissues and/or components in plasma from proinflammatory OxPL. A finding supporting this is a report that Lp(a) contains higher levels of PAF-aceylthydrolase than LDL. This enzyme has the capacity to hydrolyze and so detoxicate OxPL. A recent study from A. Tselepis reports that PAF-acetylthydrolase associated with Lp(a) in coronary artery disease patients had a reduced enzymatic effect compared to control subjects. This reduced capacity could contribute to enrichment of OxPL in Lp(a).(16)

Lp(a) is found in atherosclerotic lesions. One contributing explanation for this finding is the enhanced affinity for Lp(a) to extracellular matrix in the lesions. Some studies also claim a correlation between plasma levels of Lp(a) and quantities of Lp(a) in the lesions. As a consequence OxPL bound to Lp(a) is trapped in the lesion with the possibility to act pro-inflammatory. So, overall Lp(a) could act anti-atherogenic in the plasma compartment, but pro-atherogenic when trapped in the atherosclerotic lesion.

Conclusions: OxPL, measured as EO6 levels in apoB100 containing particles in plasma shows a strong correlation with Lp(a). Different treatments increase plasma levels of OxPL and Lp(a). The mechanisms and relationship between OxPL and Lp(a) levels are unknown, although a pattern can be noticed by the referred studies. High levels of OxPL could be considered as one measurement of oxidative stress or tissue injury, while a raise in Lp(a) levels could be part of the "acute phase response".

References:

1: Furnkranz A, Schober A, Bochkov VN, Bashtrykov P, Kronke G, Kadl A, Binder BR, Weber C, Leitinger N. Oxidized phospholipids trigger atherogenic inflammation in murine arteries. Arterioscler Thromb Vasc Biol. 2005 Mar;25(3):633-8

2: Furnkranz A, Leitinger N. Regulation of inflammatory responses by oxidized phospholipids:structure-function relationships.Curr Pharm Des. 2004;10(8):915-21.

3: Leitinger N. Oxidized phospholipids as modulators of inflammation in atherosclerosis. Curr Opin Lipidol. 2003 Oct;14(5):421-30.

4: Kagan VE, Quinn PJ. Toward oxidative lipidomics of cell signaling. Antioxid Redox Signal. 2004 Apr;6(2):199-202.

5: Binder CJ, Chang MK, Shaw PX, Miller YI, Hartvigsen K, Dewan A, Witztum JL. Innate and acquired immunity in atherogenesis. Nat Med. 2002 Nov;8(11):1218-26.

6: Chang MK, Binder CJ, Torzewski M, Witztum JL. C-reactive protein binds to both oxidized LDL and apoptotic cells through recognition of a common ligand: Phosphorylcholine of oxidized phospholipids. Proc Natl Acad Sci U S A. 2002 Oct 1;99(20):13043-8.

7: Chang MK, Bergmark C, Laurila A, Horkko S, Han KH, Friedman P, Dennis EA,Witztum JL. Monoclonal antibodies against oxidized low-density lipoprotein bind to apoptotic cells and inhibit their phagocytosis by elicited macrophages: evidencethat oxidation-specific epitopes mediate macrophage recognition. Proc Natl Acad Sci U S A. 1999 May 25;96(11):6353-8.

8: Berglund L, Ramakrishnan R. Lipoprotein(a): an elusive cardiovascular risk factor.Arterioscler Thromb Vasc Biol. 2004 Dec;24(12):2219-26

9: Edelstein C, Pfaffinger D, Hinman J, Miller E, Lipkind G, Tsimikas S, Bergmark C, Getz GS, Witztum JL, Scanu AM.Lysine-phosphatidylcholine adducts in kringle V impart unique immunological and potential pro-inflammatory properties to human apolipoprotein(a).J Biol Chem. 2003 Dec 26;278(52):52841-7

10: Bergmark C, Tsimikas S, Witztum JL. A possible novel function for lipoprotein (a) as a carrier of oxidized phospholipids. Arterioscler. Thromb. Vasc. Biol. 2004;25:e71.

11: Tsimikas S, Lau HK, Han KR, Shortal B, Miller ER, Segev A, Curtiss LK, Witztum JL, Strauss BH. Percutaneous coronary intervention results in acute increases in oxidized phospholipids and lipoprotein(a): hort-term and long-term immunologic responsesto oxidized low-ensity lipoprotein. Crculation. 2004 Jun 29;109(25):3164-70.

12: Tsimikas S, Bergmark C, Beyer RW, Patel R, Pattison J, Miller E, Juliano J,Witztum JL. Temporal increases in plasma markers of oxidized low-density lipoproteinstrongly reflect the presence of acute coronary syndromes.J Am Coll Cardiol. 2003 Feb 5;41(3):360-70.

13: Tsimikas S, Witztum JL, Miller ER, Sasiela WJ, Szarek M, Olsson AG, Schwartz GG; Myocardial Ischemia Reduction with Aggressive Cholesterol Lowering (MIRACL)High-dose atorvastatin reduces total plasma levels of oxidized phospholipids and immune complexes present on apolipoprotein B-100 in patients with acute coronary syndromes in the MIRACL trial. Circulation. 2004 Sep 14;110(11):1406-12.

14: Silaste ML, Rantala M, Alfthan G, Aro A, Witztum JL, Kesaniemi YA, Horkko S. Changes in dietary fat intake alter plasma levels of oxidized low-density lipoprotein and lipoprotein(a).Arterioscler Thromb Vasc Biol. 2004 Mar;24(3):498-503.

15: Schneider M, Witztum JL, Young SG, Ludwig EH, Miller ER, Tsimikas S, Curtiss LK, Marcovina SM, Taylor JM, Lawn RM, Innerarity TL, Pitas RE. High-level lipoprotein [a] expression in transgenic mice: evidence for oxidized phospholipids in lipoprotein [a] but not in low density lipoproteins.J Lipid Res. 2005 Apr;46(4):769-78.

16: Tsironis LD, Katsouras CS, Lourida ES, Mitsios JV, Goudevenos J, Elisaf M Tselepis AD. Reduced PAF-acetylhydrolase activity associated with Lp(a) in patients withcoronary artery disease.Atherosclerosis. 2004 Nov;177(1):193-201.

Hunting susceptibility genes for major psychoses. The TOP project

Ole Andreassen,
Ullevål University Hospital, Oslo, Norway

INTRODUCTION

The Thematic Research Area Psychotic Disorders was established at the University of Oslo in May 2003. The goal is to increase collaboration between clinical psychiatric research units, integrate clinical and basic research, maintain an infrastructure for brain imaging, molecular genetics as well as neurocognitive and clinical evaluation, and use this network to run large, multicenter research projects. Based on this collaborative effort we have recently initiated the TOP project, which purpose is to investigate the molecular genetic basis of severe mental illness by identifying distinct phenotypes based on clinical characteristics and brain function.

The TOP study is running at four psychiatric hospitals in Oslo (Aker, Ullevaal, Rikshospitalet, Diakonhjemmet/Lovisenberg). The project will combine brain imaging and molecular genetics of psychotic disorders, and is based on the advantages that Norway can offer. These are biobanks, health registries, public health care system and patients who are easily accessible and cooperative (having granted permission for their blood and tissue samples to be used for genetic research). Furthermore, together the participating hospitals are serving the whole of Oslo's population, more than 600,000 people.

The TOP project builds on, and extends, ongoing projects in psychotic disorders in Oslo funded by the Research Council of Norway (RCN) and the hospitals. The Ullevaal 600 project (PI Ole Andreassen, Ullevaal) investigates several clinical aspects of psychotic disorders (side effects, cognition, substance abuse), and includes a biobank. A first psychotic episode study (PI Ingrid Melle, Ullevaal) which is an extension of the large TIPS project, would recruit young patients. Based on a schizophrenia brain imaging study (PI Ingrid Agartz, Diakonhjemmet), a method for structural brain Magnetic Resonance Imaging (MRI) has been established. Functional MRI has been developed at Rikshospitalet during an ongoing project in affective disorders (PI Ulrik Malt). Through the different subprojects, the TOP study includes a large number of patients from all psychiatric hospitals in Oslo, integrate the structural and functional brain MRI methods already developed, and builds a large biobank. We are also in the process of developing a molecular genetic laboratory unit.

The TOP study is in line with the principles of the recent RCN evaluation of clinical research in Norway. It is based on collaboration between several research groups from different disciplines, which will be able to use resources more cost effectively and take advantage of each others knowledge. The groups are integrated with a strong leadership and the study is a result of strategic planning. The TOP study is based on international collaboration, and has moved into translational research, which is essential for gaining new knowledge in psychiatric disorders.

BACKGROUND

Severe mental disorders are characterized by the presence of psychosis, and include schizophrenia and bipolar disorder (manic-depressive disorder). They are major causes of disability worldwide and rank as some of the most costly disorders to afflict humans (Global Burden of Disease, WHO). The etiology remains elusive, but both schizophrenia and bipolar illness are now regarded as multifactorial disorders (Hyman 2000, Cowan et al. 2002). This means that both genetic and environmental factors play a role, which is the case with most

common diseases from diabetes to cancer (Zerhouni 2003). However, overwhelming evidence from adoption, twin and family studies have shown that genetic factors are most important, and the heritability for both disorders is estimated to approximately 80% (Owen et al. 2000, Cowan et al. 2002). Recently, the final sequence of the human genome was mapped. This tremendous effort unraveling the genome has opened new possibilities in the study of the neurobiology of brain by enabling us to identify genes that convey variability to brain function and human behavior, and how they interact with environmental factors in the development of psychotic disorders. The expected impact of the genomics era in psychiatry will probably be substantial (Insel & Collins 2003). Furthermore, after the sequence of the human genome was finished, an international consortium was launched that will make a haplotype map of human genome variation that is planned to be finished within 2 years (HapMap; www.genome.gov). In parallel, new powerful high-throughput genotyping techniques are being developed. This will enable the analysis of a large number of human genes as possible candidates for severe mental disorders as well as whole genome scan using haplotypes to identify regions harboring genes affecting these phenotypes (Insel & Collins 2003).

Molecular genetic research strategy in psychiatry

A new approach to the search for genetic determinants of psychiatric diseases is to look at phenotypic variables that may be more easily quantifiable, that are enduring traits rather than state dependent, and that may have a clear neurophysiological basis in that their molecular neurobiology is better understood (intermediate phenotypes). A variety of neuropsychological and neuro-physiological studies have detected abnormalities in "unaffected" family members of patients with severe mental disorders. The inherited phenotype may be subtle abnormalities of cortical function. Such phenotypes may improve the power of genetic studies by reducing genetic heterogeneity or providing phenotypes with a simpler genetic architecture. Several types of studies can be performed using such data, including association studies with candidate genes (in trios and sib pairs).

Individual variability in both normal and abnormal behavior is likely to be accounted for by polygenic systems (Quantitative Trait Loci) that result in quantitative continua of dimensions (Plomin 1999), and it is unlikely to find a single gene explaining a large fraction of the variance. In this regard, it has become increasingly accepted that genetic variation makes an important contri-bution to differences among individuals in the normal range of behavior as well as for abnormal behavior. The use of molecular genetic information along with non-invasive functional neuro-imaging has made it possible to study the underlying biological mechanisms, particularly in regard to cognitive and emotional behaviors. A simple polymorphism in the promoter region of the serotonin transporter gene can partially account for variability in the response of the amygdala to emotional stimuli (Hariri et al. 2002). Moreover, a simple polymorphism of the BDNF gene accounts for a proportion of the variance in human memory and hippocampal function (Egan et al. 2003). These studies have provided evidence that combining genetic information with in vivo brain phenotypes and with molecular techniques can be extremely useful for clarification of individual variability of higher human cognitive and emotional processes. These latter studies have also demonstrated that neuroimaging techniques such as functional MRI are incredibly powerful for detecting gene effects (Imaging Genomics). Therefore, the combination of genetic information with neuroimaging techniques, both functional and structural MRI, greatly increases statistical power thus reducing the number of subjects needed to find associations.

AIMS

The main aims of the proposed project

- Identify endophenotypes in psychotic disorders based on clinical characteristics, cognitive deficits and brain imaging
- Use the endephenotypes in the search for genetic susceptibility for the development of schizophrenia and bipolar disorder
- Study how psychopharmacological drugs and genetic factors interact in the treatment of severe mental illness.

Specific aims

1. Include a large, well characterized sample of patients with bipolar and schizophrenic disorders, and healthy controls. This will be achieved by recruiting patients from four psychiatric hospitals in Oslo, which is the clinical part of the TOP study. The goal is 1000 patients and 400 controls. The Ullevaal 600 project has already included 149 patients (May 2004). Brain imaging (MRI) investigations would be performed on 120-140 subjects per year by the TOP MRI Network.
2. Investigate if the expression of genes or groups of genes in blood cells is changed in response to psychopharmacological treatment. We would use microarray techniques to measure RNA expression of around 10000 genes to study if any genes are up or down-regulated in association with antipsychotic or mood stabilizer treatment. In particular, we intend to study a series of candidate genes already identified by in-vitro and animal experiments.
3. Use the clinical characteristics, neurocognitive function and fMRI data to identify well-defined phenotypes. The data from the clinical and MRI parts of the TOP study would be analyzed in terms of symptom cluster, disease history and severity, as well as the result on neurocognitive tests. Data from bipolar disorder and schizophrenia would be analyzed together and separately in order to identify possible endophenotype candidates.
4. Search for genetic susceptibility factors related to endophenotypes in the patient samples. We would search for SNP located in already identified candidate genes as well as in genes linked to putative pathological processes related to the disease development. High throughput SNP genotyping would be performed through the Norwegian FUGE consortium. We would use the trios to do Transmission Disequilibrium Test (TDT) and other statistical methods to calculate how the identified susceptibility genes are inherited. We would also pool our DNA with other samples abroad (Sweden, Denmark, Italy, USA).
5. Further characterize the genetic susceptibility factors identified. We would use several molecular genetic methods to characterize and verify genes discovered by SNP genotyping and RNA expression in the TOP molecular genetic laboratory.

TOP study group

The following research groups would participate in the TOP study: Clinical Psychiatric Units: Aker University Hospital, PhD Roar Fosse. Ullevaal University Hospital: Senior Scienctist Ingrid Melle, Professor Svein Friis and Professor Stein Opjordsmoen. Rikshospitalet University Hospital: Professor Ulrik Malt. Diakonhjemmet and Lovisenberg Hospital: Professor Ingrid Agartz and Senior Scientist Lars Tanum. The Neurocognitive Research Group at the Institute of Psychology with Professor Kjetil Sundet would be responsible for the neurocognitive part of the clinical studies. The gene bank (DNA extraction and long-term storage) is established at the

Norwegian Institute of Public Health, Molecular Epidemiology Laboratory (Senior Researcher Kjersti Skjold Rønningen). Through collaboration with Professor Kristian Tambs, we would intend to use the public health registries (Norwegian Birth Registry, Army Records, The Central Norwegian Person Registry) for studies of environmental factors. The fMRI studies would be performed at the fMRI group at the Rikshospitalet (Researcher Atle Bjørnerud), in collaboration with the fMRI research group at Ullevaal (Professor Per Nakstad). The Dept. of Biochemistry at Ullevaal would be responsible for the clinical biochemistry analysis (Dr. Sverre Landaas).

REFERENCE LIST

Cowan WM, et al. The human genome and its impact on psychiatry. Ann Rev Neurosci 2002, 25: 1-50.

Egan MF, Goldberg TE, Kolachana BS, et al. Effect of COMT Val 108/105 Met genotype on frontal lobe function and risk for schizophrenia. Proc Nat Acad Sci 2001; 98: 6917 – 22.

Hariri AR, Mattay VS, Tessitore A, et al. Serotonin transporter genetic variation and the response of the human amygdala. Science 2002; 297:400-3.

Hyman SE. The genetics of mental illness: implications for practice. Bull World Heal Org. 2000;78:455-63.

Insel TR, Collins FS. Psychiatry in the genomics era. Am J Psychiatry 2003:160;616-20.

Owen MJ, Cardno AG, O'Donovan MC. Psychiatric genetics: back to the future. Mol Psychiat 2000;5:22-31

Plomin R. Genetics and general cognitive ability. Nature 1999, 402: C25-9.

Zerhouni, E. The NIH Roadmap. Science 2003; 302:63-72

Psychiatric genetics: A current perspective
Part I – Genetic epidemiology
Part II – Gene finding methods

Kenneth Kendler
VCU Life Science Faculty, USA

The talks will begin with an elaboration of the four major paradigms in psychiatric genetics: simple genetic epidemiology, advanced genetic epidemiology, gene-finding methods, and molecular biology. The strengths and limitations of each of these methods will be outlined, as well as their inter-relationships.

Then, over the course of the two talks, pertinent illustrations will be provided for the first three paradigms. We will review the heritability of the major psychiatric disorders. We will then review a selected set of findings from advanced genetic epidemiology. These will include: modeling genotype-environment interaction, genotype-environment correlation, genetic epidemiologic models of development, gene by cohort interactions and gender effects. Examples will be provided of integrated etiologic models incorporating multiple environmental risk factors into genetically informative designs.

The section on gene-finding methods will begin with a brief review of association and linkage methodologies as applied to complex traits. We will discuss the conceptual issue of what it means to have a "gene for" a disorder. This discussion will conclude that such phraseology, although commonly used in the professional and lay literature, is probably inappropriate for genetic influences on psychiatric illness.

We will then review in some detail the efforts to map—using linkage analysis—susceptibility genes for schizophrenia; and follow through the story on one potential positional candidate gene, Dysbindin. This will lead us into questions of what the unit of replication is in these analyses, that is: the allele at a marker, that individual marker, the haplotype or the gene itself. The current status of the findings in Dysbindin will be discussed and the problems with the interpretation of the heterogeneous results.

We will then review a second linkage field in the area of alcohol dependence; and then describe some consistency of findings across studies on chromosome *4q*.

These discussions will conclude with a broad overview of how the field of psychiatric genetics is likely to progress in subsequent years.