



HAL
open science

Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association.

Françoise Clerget-Darpoux, Robert C. Elston

► To cite this version:

Françoise Clerget-Darpoux, Robert C. Elston. Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association.. Human Heredity, Karger, 2007, 64 (2), pp.91-6. 10.1159/000101960 . inserm-00129091

HAL Id: inserm-00129091

<https://www.hal.inserm.fr/inserm-00129091>

Submitted on 4 Sep 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Invited editorial

Is linkage and the collection of family data dead? Prospects for family studies in the age of genome-wide association

Françoise Clerget-Darpoux^{1,2} and Robert C. Elston³

¹INSERM U535, IFR 69, Villejuif, France

²Université Paris-Sud, Villejuif, France

³Case Western Reserve University, Cleveland, Ohio, U.S.A.

The HapMap project and the discovery of millions of single nucleotide polymorphisms (SNPs) throughout the human genome, together with the development of bead and chip technology, has made it feasible to type hundreds of thousands of markers on an individual for a few hundred dollars. Because of the extensive linkage disequilibrium (LD) that exists, genome-wide association studies are currently being conducted on samples of unrelated persons in the belief by some that this is now the design of choice to discover genetic variants underlying relatively common complex diseases. Huge tissue repositories have been set up for the purpose of conducting large-sample case-control studies. Should we therefore stop collecting family data, forget that we inherit our genes from our parents and ignore the fundamental laws of genetic transmission as being unnecessary for gene discovery? Can we hope to anticipate, understand and treat the many diseases to which humans are prone, simply by finding genomic locations that differ between those who have and those who do not have disease? Below we argue that, although they have a useful role to play, it should not be naively assumed that the time has come for genome-wide association studies to replace genome-wide linkage studies, or that the collection of family data is no longer necessary.

The power to detect a gene

Risch and Merikangas (1996) wrote “*We argue that {the linkage} method has limited power to detect genes of modest effect*”[1].

What is meant by a gene of modest effect? A classical measure of the genetic effect of a diallelic locus on a binary trait is the odds ratio (OR), either for one versus no copies of the susceptibility variant allele, or for two versus one copy of it, these two measures being equal under a multiplicative model. It is important to realize that, in an association study with equal numbers of cases and controls, the power to detect such a variant depends only on the ORs; whereas in a linkage study it depends on both the ORs and the variant allele frequencies in the sample. From the data of Bell et al. [2] on type I diabetes in the North American population, for example, the class 1 allele (M1) of the variable number of tandem repeats (VNTR) flanking the Insulin gene has a frequency of 0.70 and we can estimate under a multiplicative model an OR of 3.4 for this allele. There is a good power to detect this effect in a case-control association study, but little power to detect it by linkage in affected sib-pairs. On the other hand, had the allele frequency been 0.10, the same OR would lead to an affected sib pair (ASP) linkage study with very good power.

When a genome-wide scan is performed and only the most significant results noted, allowance must be made for multiple testing, and the effect this has on power is quite different for linkage and association studies because of the enormous difference in the effective number of independent tests that each entails. If, on the other hand, there is good motivation for examining a candidate gene or candidate region (detected by a previous linkage analysis), an association study becomes the second stage of a multiple sampling strategy, which may lead to a much more economical study to obtain a given amount of power [3] Linkage analysis can similarly detect what interactions may exist [4] and in this way also limit the number of interaction models to examine in an association analysis.

The problem of replication

“Replication has a vital role in showing that associations that are identified reflect interesting biological processes rather methodological quirks” [5].

Once an initial study has detected a susceptibility variant, heterogeneity from population to population with respect to either environmental or genetic factors can make it difficult to obtain an appropriate sample to perform a replication study. First, the replication sample will need to be as large or even larger [6;7], and second, the difference in both variant frequencies and LD patterns among populations may make the signal undetectable in other populations. The difficulty of replicating an association study is well illustrated in the work of Reich et al [8] on Multiple Sclerosis, where an association on chromosome 1 in African-Americans was not replicated in a different sample of Afro-Caribbeans. Similarly, an association between Myosin IXB (MYO9B) gene variants and celiac disease (CD) was recently detected in a Dutch case-control association study [9], but not confirmed in the British [10], Swedish/Norwegian [11], Italian [12] or Spanish [13] populations. In these studies, we cannot determine whether the first reported association was driven by a causal variant in the one population and not in the others, or whether it was just a false positive result. Whereas Lohmueller et al [14] argue that a sizable fraction of reported genetic associations have strong evidence of replication, it is well established that association studies in general have a high tendency not to be replicated, the first report being simply a chance occurrence [15]

What are the factors determining disease susceptibility?

Common versus rare variants

The poor success of genome-wide linkage searches for complex diseases has imposed in a hegemonic way the “common disease-common variant” (CDCV) hypothesis, which is the major motivation for the HAPMAP project. By their very construction, common tagging SNPs do not identify rare variants, yet we know that multiple rare variants in the same gene can play an important role in disease susceptibility [16-18]. The total of all the susceptibility allele frequencies

may be quite high, but with extensive allelic heterogeneity. This is the case, for example, in Crohn's disease - for which risk alleles in the NOD2 gene have frequencies each smaller than 5% in populations of European descent [19;20], but with great variation of the total risk allele frequencies among those populations. Models of the expected genetic variation at disease susceptibility loci based on population genetics have suggested that when selection is weak, as seems likely for many complex disease mutations, genetic drift becomes important and the total frequency of susceptibility mutations is expected to vary widely among loci. Equilibrium models and empirical studies suggest a role for both rare and common variants [18]. Indeed, for classic monogenic diseases allelic heterogeneity is the rule rather than the exception: is there any reason to think this should be different for complex disease?

Gene-gene and gene-environment interaction

In the case of a quantitative trait, the presence of interaction (non-additivity) can depend on the scale of measurement. Tukey's test for non-additivity [21] specifically detects removable non-additivity, i.e. non-additivity that can be removed by application of a monotonic transformation to the data [22]. Recently, Chatterjee et al [23] have used a quantitative score based on SNPs as an indicator of a binary trait and hence have been able to use Tukey's idea to increase the power of testing main effects in the case of a binary trait when removable interaction exists; but non-removable interaction is not detected. The interactive effect of several functional genetic variants may play a larger role than the marginal effect of each variant in determining a trait, especially in the case of multifactorial diseases. Non-removable interactions of this sort cannot be easily found in genome-wide studies without very large samples. On the other hand, if we restrict our search for interactions by only considering simultaneously genes that belong to a relevant pathway, or to the *a priori* most likely types of joint gene action, correspondingly restricted association studies become feasible. Of course, it must always be remembered that "a gene association study that begins by

looking at *a priori* functional candidates is looking on average for a different class of variants than is a positional cloning study that started from a significant linkage signal”[16]. Nevertheless, as we noted above, prior information obtained from linkage studies can be used to motivate a more limited number of interactions to consider when conducting an association analysis.

In autoimmune diseases, the HLA component acts through complex interactions of numerous variants. A deeper understanding of these diseases can only be obtained through more detailed knowledge of gene diversity, patient genotype distributions, IBD sharing in affected siblings conditional on patient genotype, and a biological understanding of the function of the genes in the HLA region. Detecting an association gives us very limited information, especially when we wish to estimate risk, to which we now turn.

Estimation of risk

It is a far cry from detecting a signal indicating the presence of a causative factor in a genomic region to its identification and the clinically important task of estimating the disease risk due to it. In the first place, the same population distribution of a marker can be coupled with different modes of inheritance of the trait, and hence different risk estimates. For example, the observed distribution of the VNTR flanking the Insulin gene (coded as a diallelic marker: M1 for less than 1000 repeats, M2 for a larger repeat number) in Type I Diabetes [2] may be the result of either

- a direct role of the VNTR (**situation 1**),

in which case the allele frequencies are (0.67, 0.33) and the genotypic relative risks of M1M2 and M2M2 versus the risk of M1M1 are $GRR_1 = 0.33$ and $GRR_2 = 0.06$, respectively; or

- a causal diallelic variant (S1, S2) in LD with the VNTR (M1, M2), with $P(S1/M1) = 0.2$, $P(S2/M1) = 0.8$ and $P(S1/M2) = 1$ (**situation 2**).

In this case the causal variant has very different allele frequencies (0.13, 0.87) and the genotypic relative risks of S1S2 and S2S2 versus the risk of S1S1 are $GRR_1 = 0.10$ and $GRR_2 = 0.004$, respectively.

We can distinguish between these two situations if we know the allele sharing probabilities for the insulin gene among ASPs, which for these two cases, letting z_i denote the probability that they share i alleles IBD, are respectively:

$$(z_2, z_1, z_0) = (0.29, 0.50, 0.21) \quad \text{in situation 1}$$

and $(z_2, z_1, z_0) = (0.56, 0.39, 0.05) \quad \text{in situation 2.}$

In fact, on the basis of 95 such sib pairs [24], we know that the former is the true situation, i.e. the VNTR has a direct role, as has been confirmed by functional studies. Thus linkage information is crucial in order to discriminate between these two situations that lead to the same association signal. It is important to identify models that can explain both the association and the linkage information that is found, and not just one of these [25], if we are to obtain good estimates of risk. It was also the IBD sharing information on sibs affected with Rheumatoid Arthritis which showed that the usual classification [26] of the DRB1 risk alleles was incorrect because it neglected numerous infrequent risk alleles [27].

Moreover, note that linkage information, far from being redundant once an association is found, can be confirmatory. If we have a sample of affected sib pairs, information derives from the IBD distribution of an affected sibling conditional on that of the index sib [28]. When the index patients have a high risk genotype, their affected sibs share two alleles IBD with a probability greater than 0.25 and, conversely, when the index patients have a low risk genotype, their affected sibs share two alleles IBD with probability less than 0.25. One example of this is the above cited Insulin VNTR, for which the overall IBD sharing is close to the null hypothesis values (0.25, 0.50, 0.25). When the index case is homozygous for the susceptibility allele, the proportions of affected sibs sharing two, one and zero alleles IBD are 0.27, 0.60 and 0.13, respectively. When the index cases have another genotype, the proportions are 0.07, 0.53, 0.40, respectively [29]. A similar situation holds for the PTP22 gene in Rheumatoid Arthritis. Four different genotypic risk groups can be defined and the proportions of ASPs sharing 2 alleles IBD are 0.30, 0.24, 0.22, 0.09 according to the risk group (from the highest to the lowest) that the index sib belongs to (Bourgey et al, personal communication). Thus linkage data contain considerable information

capable of aiding in the discrimination of causal variation from merely nearby variation that is in LD with a causal variant.

In a recent study, Lincoln et al [30] used a large-scale SNP association strategy to study the involvement of the major histocompatibility complex in Multiple Sclerosis (MS). The strongest association was observed with haplotype blocks in the HLA class II region and, conditional on this association, there was no evidence for other HLA-region risk factors for MS. Despite the use of a large number of simplex and multiplex families with dense SNP typing, the study simply concluded with a result that has been known for many years. This is another illustration of the limitations of strategies that rely solely on association.

The further importance of family information

As we have just seen in the above paragraph, collecting family information is important to refine our genetic model and risk estimates. For both clinicians and the relatives of a patient, we need estimates of familial risk. For example, parents of a child affected with celiac disease (CD) often ask for the risk to a future child. Thanks to good epidemiological familial surveys conducted in Italy, the risk that will be given to Italian parents is about 10%. It is now possible to provide even more accurate information by performing HLA-DQ typing in the family. The risk, in fact, ranges from 1 to 30%: in some situations one will be able to reassure the parents with a risk of less than 1%, whereas if a baby is at high risk (about 30%) specific follow-up may be offered [31].

A major source of complexity in disease etiology is underlying heterogeneity. In particular, it is possible for a disease that initially appears to be multifactorial to have monogenic sub-entities. Detection of such heterogeneity will be impossible through association studies, but is enabled by family studies. This is well illustrated by breast cancer, for which the existence of monogenic forms (crucial information in terms of public health) was revealed by segregation analysis in a very large sample of families containing affected women [32]. The location of BRCA1 [33], as well as its identification [34] was only possible through familial studies. By taking into account the family history of a woman a

search of a particular mutant allele may be indicated for her medical follow-up. Similar examples could be given for other types of cancer, for neurological diseases (e.g. Alzheimer disease, frontal dementia, Parkinson disease), for cardiovascular diseases (cardiomyopathy), etc.

Note, too, that the well-established phenomena imprinting and anticipation can only be detected if family data are collected. Similarly, there have been reports of maternal genotype effects for neural tube defects [35;36] and autism [37]. These and epigenetic phenomena leading to transgenerational effects [38;39] can only be detected if family data are collected. Finally, one could note that it is much easier to detect typing errors in family data – though this may be of less importance in view of the high accuracy of the current platforms available for genome-wide association studies provided appropriate steps are taken to detect errors that are known to occur, such as because of abnormal copy number.

Conclusion

Finally, let us return to the quotation above from Risch and Merikangas [1], which has often been quoted to justify genome-wide case-control association studies (though we note in passing that the comparison they made was between an ASP linkage study and a study jointly testing linkage and association, i.e. performing genome-wide transmission disequilibrium tests - TDTs) and take it two sentences further: “*We argue that {the linkage} method has limited power to detect genes of modest effect. A different approach that utilizes **candidate genes has far greater power**, even if one needs to test every gene in the genome. Thus, the future of the genetics of complex diseases is likely to require large-scale testing by **association analysis**” [emphasis ours].*

Note carefully it is erroneous to assume that a candidate gene study cannot be conducted using linkage analysis, and we have seen above that association information should not be considered as being in some way superior to linkage

information. We firmly believe that the current efforts being put into the construction of huge databases for case-control studies should not be done to the detriment of continued collection of family data. So many different biological mechanisms are possible that it would be foolhardy to restrict all human genetic research to a single strategy. This could well lead to the same frustrations that followed the wholesale adoption of genome-wide linkage scans (Google counts of these genome-wide studies so far: 183,000 for linkage, 181,000 for association). One of the greatest current challenges facing human genetics is that of how best to gather and synthesize the many lines of evidence possible in order to discover the genetic determinants underlying complex diseases. Just as previously, before we had so many genetic markers available, it was well recognized that it was necessary to use many different lines of evidence to confirm a genetic component in the etiology of a disease – population studies, adoption studies, twin and family studies - so now, with the availability of all these genetic markers, must we diversify the kinds of studies we perform when using this new tool for the purpose of dissecting these genetic components for the betterment of mankind. A wise investigator, just like a wise investor, should hedge all bets.

Acknowledgements

This work was supported in part by a U.S. Public Health Research grant (GM28356) from the National Institute of General Medical Sciences and Cancer Center Support Grant (P30CAD43703) from the National Cancer Institute.

References

- 1 Risch N, Merikangas K: The future of genetic studies of complex human diseases. *Science* 1996;273:1516-1517.
- 2 Bell GI, Horita S, Karam JH: A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 1984;33:176-183.

- 3 Elston RC, Lin D, Zheng G: Multi-stage sampling for genetic studies. *Annu Rev Genomics Hum Genet* 2007;in press.
- 4 Gauderman WJ, Morrison JL, Siegmund KD: Should we consider gene x environment interaction in the hunt for quantitative trait loci? *Genet Epidemiol* 2001;21 Suppl 1:S831-S836.
- 5 Hattersley AT, McCarthy MI: What makes a good genetic association study? *Lancet* 2005;366:1315-1323.
- 6 Terwilliger JD, Haghghi F, Hiekkalinna TS, Goring HH: A bias-ed assessment of the use of SNPs in human complex traits. *Curr Opin Genet Dev* 2002;12:726-734.
- 7 Sohn D: Statistical significance and replicability: Why the former does not presage the latter. *Theory Psychology* 1998;8:291-311.
- 8 Reich D, Patterson N, De Jager PL, McDonald GJ, Waliszewska A, Tandon A, Lincoln RR, DeLoa C, Fruhan SA, Cabre P, Bera O, Semana G, Kelly MA, Francis DA, Ardlie K, Khan O, Cree BA, Hauser SL, Oksenberg JR, Hafler DA: A whole-genome admixture scan finds a candidate locus for multiple sclerosis susceptibility. *Nat Genet* 2005;37:1113-1118.
- 9 Monsuur AJ, de Bakker PI, Alizadeh BZ, Zhernakova A, Bevova MR, Strengman E, Franke L, van't SR, van Belzen MJ, Lavrijsen IC, Diosdado B, Daly MJ, Mulder CJ, Mearin ML, Meijer JW, Meijer GA, van OE, Wapenaar MC, Koeleman BP, Wijmenga C: Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. *Nat Genet* 2005;37:1341-1344.
- 10 Hunt KA, Monsuur AJ, McArdle WL, Kumar PJ, Travis SP, Walters JR, Jewell DP, Strachan DP, Playford RJ, Wijmenga C, van Heel DA: Lack of association of MYO9B genetic variants with coeliac disease in a British cohort. *Gut* 2006;55:969-972.
- 11 Amundsen SS, Monsuur AJ, Wapenaar MC, Lie BA, Ek J, Gudjonsdottir AH, Ascher H, Wijmenga C, Sollid LM: Association analysis of MYO9B gene polymorphisms with celiac disease in a Swedish/Norwegian cohort. *Hum Immunol* 2006;67:341-345.
- 12 Giordano M, Marano C, Mellai M, Limongelli MG, Bolognesi E, Clerget-Darpoux F, Momigliano-Richiardi P, Greco L: A family-based study does not confirm the association of MYO9B with celiac disease in the Italian population. *Genes Immun* 2006;7:606-608.
- 13 Nunez C, Marquez A, Varade J, Martinez A, Polanco I, Maluenda C, Fernandez-Arquero M, de la Concha EG, Urcelay E: No evidence of

association of the MYO9B polymorphisms with celiac disease in the Spanish population. *Tissue Antigens* 2006;68:489-492.

- 14 Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN: Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003;33:177-182.
- 15 Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG: Replication validity of genetic association studies. *Nat Genet* 2001;29:306-309.
- 16 Pritchard JK, Cox NJ: The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet* 2002;11:2417-2423.
- 17 Pritchard JK: Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001;69:124-137.
- 18 di Rienzo A: Population genetics models of common diseases. *Curr Opin Genet Dev* 2006;16:630-636.
- 19 Lesage S, Zouali H, Cezard JP, Colombel JF, Belaiche J, Almer S, Tysk C, O'Morain C, Gassull M, Binder V, Finkel Y, Modigliani R, Gower-Rousseau C, Macry J, Merlin F, Chamaillard M, Jannot AS, Thomas G, Hugot JP: CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002;70:845-857.
- 20 Vermeire S, Wild G, Kocher K, Cousineau J, Dufresne L, Bitton A, Langelier D, Pare P, Lapointe G, Cohen A, Daly MJ, Rioux JD: CARD15 genetic variation in a Quebec population: prevalence, genotype-phenotype relationship, and haplotype structure. *Am J Hum Genet* 2002;71:74-83.
- 21 Tukey JW: One degree of freedom for non-additivity. *Biometrics* 1949;5:232-242.
- 22 Elston RC: Additivity in the analysis of variance. *Biometrics* 1961;17:209-219.
- 23 Chatterjee N, Kalaylioglu Z, Moslehi R, Peters U, Wacholder S: Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. *Am J Hum Genet* 2006;79:1002-1016.
- 24 Spielman RS, Baur MP, Clerget-Darpoux F: Genetic analysis of IDDM: summary of GAW5 IDDM results. *Genet Epidemiol* 1989;6:43-58.

- 25 Clerget-Darpoux F, Babron MC, Prum B, Lathrop GM, Deschamps I, Hors J: A new method to test genetic models in HLA associated diseases: the MASC method. *Ann Hum Genet* 1988;52:247-258.
- 26 Gao X, Gazit E, Livneh A, Stastny P: Rheumatoid arthritis in Israeli Jews: shared sequences in the third hypervariable region of DRB1 alleles are associated with susceptibility. *J Rheumatol* 1991;18:801-803.
- 27 du Montcel ST, Michou L, Petit-Teixeira E, Osorio J, Lemaire I, Lasbleiz S, Pierlot C, Quillet P, Bardin T, Prum B, Cornelis F, Clerget-Darpoux F: New classification of HLA-DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. *Arthritis Rheum* 2005;52:1063-1068.
- 28 Clerget-Darpoux F, Babron MC, Bickeboller H: Comparing the power of linkage detection by the transmission disequilibrium test and the identity-by-descent test. *Genet Epidemiol* 1995;12:583-588.
- 29 Dizier MH, Babron MC, Clerget-Darpoux F: Interactive effect of two candidate genes in a disease: extension of the marker-association-segregation chi(2) method. *Am J Hum Genet* 1994;55:1042-1049.
- 30 Lincoln MR, Montpetit A, Cader MZ, Saarela J, Dymont DA, Tiislar M, Ferretti V, Tienari PJ, Sadovnick AD, Peltonen L, Ebers GC, Hudson TJ: A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. *Nat Genet* 2005;37:1108-1112.
- 31 Bourgey M, Calcagno G, Tinto N, Gennarelli D, Margaritte-Jeannin P, Greco L, Limongelli M-G, Esposito O, Marano C, Troncione R, Spanpanato A, Clerget-Darpoux F, Sacchetti L: Genetic Counseling in Celiac Disease. *Gut*. In press.
- 32 Newman B, Austin MA, Lee M, King MC: Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. *Proc Natl Acad Sci U S A* 1988;85:3044-3048.
- 33 Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC: Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* 1990;250:1684-1689.
- 34 Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W, .: A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 1994;266:66-71.
- 35 Doolin MT, Barbaux S, McDonnell M, Hoess K, Whitehead AS, Mitchell LE: Maternal genetic effects, exerted by genes involved in homocysteine remethylation, influence the risk of spina bifida. *Am J Hum Genet* 2002;71:1222-1226.

- 36 Brody LC, Conley M, Cox C, Kirke PN, McKeever MP, Mills JL, Molloy AM, O'Leary VB, Parle-McDermott A, Scott JM, Swanson DA: A polymorphism, R653Q, in the trifunctional enzyme methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase/formyltetrahydrofolate synthetase is a maternal genetic risk factor for neural tube defects: report of the Birth Defects Research Group. *Am J Hum Genet* 2002;71:1207-1215.
- 37 Jones MB, Palmour RM, Zwaigenbaum L, Szatmari P: Modifier effects in autism at the MAO-A and DBH loci. *Am J Med Genet B Neuropsychiatr Genet* 2004;126:58-65.
- 38 Li S, Hursting SD, Davis BJ, McLachlan JA, Barrett JC: Environmental exposure, DNA methylation, and gene regulation: lessons from diethylstilbesterol-induced cancers. *Ann N Y Acad Sci* 2003;983:161-169.
- 39 Ruden DM, Xiao L, Garfinkel MD, Lu X: Hsp90 and environmental impacts on epigenetic states: a model for the trans-generational effects of diethylstilbesterol on uterine development and cancer. *Hum Mol Genet* 2005;14 Spec No 1:R149-R155.