SMP: Epigenetics

Project: DNA Methylation Profiling of Human Chromosome 21

Prof. Dr. Albert Jeltsch, International University Bremen, a.jeltsch@iu-bremen.de
Dr. Richard Reinhardt, Max-Planck-Institute for Molecular Genetic (MPI-MG), Berlin
Dr. Matthias Platzer, Institute of Molecular Biotechnology, Jena
Prof. Dr. Jörn Walter, Saarland University, Saarbrücken

Introduction

DNA methylation represents "heritable" epigenetic information in the genome [Jones & Takai, 2001; Bird, 2002]. Methyl groups are attached to cytosines in the genome and this information can be copied during DNA replication. The modification changes the recognition of the DNA content but not the DNA sequence itself. DNA methylation is a major source of regulatory information encoded on the DNA and controls a variety of genomic functions including the developmental and tissue specific expression of genes. Methylation occurs predominantly at CG sites, which are modified to a level of 70-80%. However, the distribution of methylated CpGs is not random and highly regulated during development and differentiation. Most parts of the human genome are strongly depleted in CpG sequences and most CpG dinuleotides cluster in CpG islands (on average 200-500 base pairs), occurring in about 50% of all promoters and first exon sequences of genes. Although it is generally believed that CpGs in promoter islands remain unmethylated in all developmental stages some ex-hibit a tissue and developmental specific methylation pattern. As a general rule, methylation of the promoter regions of genes leads to a strong reduction in gene expression. The genomic distribution of DNA-methylation in different tissues and cells needs to be deciphered since changes in DNA methylation have important clinical relevance: erroneous methylation is a major cause of cancer [Jones & Baylin, 2002]. It is also implicated in brain function, aging and development and differences in the epigenetic state of the DNA including methyla-tion changes are likely to be involved in the phenotype of many multifactorial diseases [Reik et al., 2001; Jones & Takai, 2001; Jaenisch & Bird, 2003]. Consequently, the detection of methylation changes is an important diagnostic approach. However, since the differences and heterogeneities of the methylation patterns between different cell types, different developmental stages and different individuals are poorly understood today, com-prehensive epigenomic maps are urgently required.

Obtaining the complete epigenomic maps of a human being is a challenging task for the future, given the current state of technology, in which most of the experimental steps including bisulphite conversion, PCR and sequencing are only partially suitable for automated platforms. Ideally, the epigenomic map should be available for the whole genome of as many as possible of the approx. 200 human cell types at different states of development. As an entry into this Human Epigenome Project, we plan to analyse the methylation of promotors and CG rich regions within genes on chromosome 21 and establish the first comprehensive epigenomic map of an entire chromosome. We expect to see methylation differences in the 284 annotated genes reflecting expression levels and tissue specificity. We have chosen Chromosome 21 since it is a particularly attractive target for such a pilot project: It contains a manageable number of well annotated genes (promoters), and the German DHGP and NGFN-1 made substantial contributions to its sequence analysis [Hattori et al., 2000]. Furthermore, chromosome 21 has been subject to many interesting comparative studies [Fujiyama et al, 2001, Gitton et al., 2002], particularly to primates and it is linked to gene dosage effects (trisomies) which are of particular relevance for disease. With our study, we will establish the basis to address the following fundamental biological and medical questions: 1. how do different methylation patterns within the promoters of human genes look, 2. how many genes show DNA-methylation variation and can this variation be linked to gene expression, 3. how much do epigenomic maps vary between individuals, 4. are genetically identical monozygotic twins different in their epigenome profiles, 5. do trisomies lead to characteristic epigenetic changes and 6. is the expression of disease genes located on chromosome 21 influenced by DNA methylation.

In general, the results of our project will provide detailed views on the pattern of DNA methylation in human cells on a representative chromosome. It will complement to an ongoing Epigenome project led by Epigenomics and the Sanger Institute. Thereby, it will provide essential information for future work in this field.

Planned work

We plan to investigate the DNA methylation pattern of Chromosome 21 from six pairs of twins (three male and three female). This will allow us to answer some very fundamental questions, on the relationship between genetic and epigenetic information, the stability of the epigenetic setting and the difference between male and female patterns. It is essential to study the methylation pattern of a homogenous population of cells, because different cells may contain different methylation information. Therefore, analysis of a mixture of cell types does not allow drawing conclusions at the level of individual cells, which is necessary for functional interpretation. Since the biological material should be readily available, we decided to focus on blood monocyts, which can be purified >95% using CD14. We are prepared to investigate other tissues (liver, lung, brain, tumors etc.) in cooperation with one of the KG networks of the NGFN-2 in the second phase of the project; a cooperation with the Cancer Network is planned. Blood samples from twins will be purchased from HealthTwist. Purification of monocyts is established in many labs of the NGFN network, e.g. in the medical department of the University of Giessen, at the GBF Braunschweig, the Rheumaforschungszentrum (Prof. A. Radbruch) in Berlin and MPI Bremen (Prof. R. Amann). FACS or MACS sorting will be performed in cooperation with some of these groups. Biochemical experiments including, DNA preparation, Bisulfite conversion of the DNA and PCR will be performed in the labs in Saarbrücken, Bremen, which have strong experience with the technique. All labs will regularly exchange samples and compare the results obtained to ensure a high technical standard of all steps and comparability of the results.

There are 284 manually annotated genes on Chromosome 21. Detailed information is available for many promoters and more information is expected to come out in near future. This knowledge represents an important and essential starting point of this project, since the design of suitable PCR-products is not trivial and cannot make use of available primer design programs. Technically, the project will require the bisulfite treatment of the DNA, which converts cytosine residues into uracil but leaves methylated cytosine residues intact. Thereby, the epigenetic information of methylation is converted into a stable genetic form. Converted DNA will be

amplified in segments of about 200-400 bps by conventional PCR ("amplicons"). PCR amplicons will be sequenced directly using conventional terminator dyes and automated capillary systems. In a first analysis, those amplicons that are almost unmethylated will be identified and excluded from the further studies. For the methylated amplicons, direct sequencing as well as cloning of the converted DNA and sequencing of individual clones will be carried out. The relative proportion of both methods will depend on the complexity of the methylation patterns (complex patterns can only be detected by sequencing of individual clones) and the success rate and reliability of direct sequencing. In addition the number of methylated amplicons that are subject to a second more detailed analysis will influence the proportion of direct sequencing and cloning. We estimate that including failures, each amplicon will be sequenced 5 times on average.

It is one limitation of the current state of the Bisulfite technology that only relatively short products can be successfully amplified from bisulfite treated DNA (typically 300-400 base pairs), because the bisulfite treatment conditions inevitably lead to fragmentation of the DNA. Therefore, we will prepare 4 to 5 independent PCR products for each gene/promoter covering about 1500-2000 base pairs around the transcriptional start point and within CpG rich regions. This approach will require approximately 1250 different PCR products to be produced.

Bisulfite converted DNA is of very low sequence complexity, which complicates design of primers suitable for amplification. These primers need to fulfil several conditions: i) only bisulfite converted DNA should be selectively amplified, ii) amplification should be independent on the methylation state and iii) the primer should be suited for automatic DNA sequencing. Therefore, efficient primer design strategies have to be developed, the PCR reactions have to be optimized and the sequencing protocols may need some adaptation. This is particularly important for the data analysis, because in this project the amount of two bases at one position (T reflecting unmethylated C before bisulfite treatment and C reflecting methylated C) have to be detected and quantified. The chromatograms have to be analysed automatically with respect to the relative amount of C and T at each CG site and also preevaluated with respect to technical quality and completeness of bisulfite conversion. In addition, manual inspection may become necessary.

In addition, the quantitative interpretation of the direct sequencing reactions has to be regularly validated on selected amplicons by independent methods. To this end, arbitrarily chosen bisulphite-PCR products will be cloned and individual clones will be sequenced. The combined methylation profiles will be compared to the results of the direct sequencing. In total about 75000 sequencing reactions have to be performed.

Given that 1250 PCR products are scheduled for analysis and 12 different DNA samples are to be studied, 15000 PCR products have to be delivered, 7500 each by the labs in Saarbrücken and Bremen. This includes FACS sorting, DNA preparation, bisulfite treatment, primer design, PCR and quality checks. In addition, regular calibration experiments and double blind analyses must be performed. Finally the data must be interpreted in the context of the expression analysis data and prepared for publication. After an initial phase of training, setting up the systems and quality checks (about 4-6 months), we estimate an annual output of 3000 PCR products from each academic lab. Data analysis, storage and data interpretation will be performed in a collaborative effort between the participating labs in Bremen, Saarbrücken, the sequence-bioinformatics of MPI, Epigenomics AG and the Sanger Institute, UK. The data will be made publicly available.

The project will be integrated with a European initiative led by Epigenomics and the Wellcome Trust Sanger Institute, which has investigated the methylation of Chromosome 6 (http://www.epigenome.org/). We will stick to the quality criteria of this project (or outreach it) to guarantee comparability of the data sets. Epigenomics will provide existing know-how in data analysis and bioinformatics and integrate the work in this project with the ongoing Epigenome project to avoid overlap and duplication of work, in close collaboration with the Sanger Institute. Data will be made available to Epigenomics on a timely basis. The results will be imported also into GenomeMatrix, a gene-centered visualization tool developed by MPI-MG in NGFN1 for the display of complex data sets from different sources..

Lit.: **1.** Bird, A. (2002) Genes Dev 16, 6-21. Jones, P. A. & Takai, D (2001). Science 293, 1068-70. **2.** Jones, P. A. & Baylin, S. B. (2002) Nat. Rev. Genet. 3, 415-28. **3.** Jaenisch, R. & Bird, A. (2003) Nat Genet 33, 245-54. **4.** Reik W, Dean W, Walter J (2001) Science 293, 1089-1093. **5.** Hattori et al., (2000) Nature 405, 311-9. **6.** The International Human Genome Mapping Consortium (including: Reinhardt, R. and Lehrach, H.) (2001) Nature 409, 934-941. **7.** Fujiyama A, et al. (2002) Science. 295, 131-4. **8.** Gitton Y et al. (2002) Nature. 420, 586-90.