RESTRICTION SITE TAGGED PASSPORTS AND MICROARRAYS FOR ANALYSIS OF COMPLEX BIOLOGICAL SYSTEMS

Zabarovsky E.R.*, 1, Kashuba V.I., 1 Li J., 1 Kutsenko A.S., 1, Protopopov A.I., 1 Petrenko L., 1 Wang E., 1 Senchenko V.N., 3 Kadirova E., 4 Zabarovska V.I., 1 Muravenko O.V., 2, Zelenin A.V., 2, Kisselev L.L., 2 Winberg G., 1, Ernberg I., 1 Braga E., 4, Lerman M.I., 5 Klein G.

1 MTC and CGB, Karolinska Institute, Stockholm, Sweden; 2 EIMB, RAS, Moscow, Russia; 3 Bioengineering Center, RAS, Moscow, Russia; 4 Russian State Genetics Center, Moscow, Russia; 5 FCI-NCI, Frederick, USA

* Corresponding author: e-mail: eugzab@ki.se

Keywords: genome scanning, CpG islands, NotI sites, epigenetics, tumor suppressor genes, methylation, genomic microarrays, biodiversity, NotI-tags, human gut

Summary

Motivation: The main aim is to develop methods (experimental and computational) for the fast analysis of prokaryotic and eukaryotic genomes and complex biological systems like human gut microbial flora and cancer cells.

Results: We have developed novel tools for genome analysis (patent pending): restriction site tagged (RST) microarrays and restriction site tagged sequences (RSTS or passporting). Using NotI enzyme we have shown that NotI microarrays offer a powerful tool with which to study carcinogenesis. Moreover, NotI microarrays are the only existing microarrays giving the opportunity to detect simultaneously and differentially copy number and methylation changes. Thus they allow to check cancer cells for genetic and epigenetic abnormalities. For microbial identification NotI microarrays and passporting are significantly more specific and powerful than sequencing of 16S ribosomal genes or any gene specific microarrays. They allowed identification in bacterial mixtures thousands of known and novel microbial species and strains. Pilot experiment showed that human intestine contains more than 90 % of unknown bacteria.

Introduction

The presence of genetic alterations in tumors is widely accepted, and explains the irreversible nature of tumors. However, now, DNA methylation in CpG sites is known to be precisely regulated in tissue differentiation, and is supposed to be playing a key role in the control of gene expression in mammalian cells. The genes involved include tumor suppressor genes, genes that suppress metastasis and angiogenesis, and genes that repair DNA suggesting that epigenetics plays an important role in tumorigenesis. It becomes clear that methylation is a basic, vital feature/mechanism in mammalian cells. It is involved in hereditary and somatic cancers, hereditary and somatic diseases, apoptosis, replication, etc. It is suggested that it can be used for diagnostic, prognostic, prediction and even for direct treatment of cancer. Based on the growing understanding of the roles of DNA methylation, several new methodologies were developed to make a genome-wide search for changes in DNA methylation: restriction landmark genomic scanning (RLGS), methylation-sensitive-representational difference analysis (MS-RDA), methylation-specific AP-PCR (MS-AP-PCR), methyl-CpG binding domain column/segregation of partly melted molecules (MBD/SPM) and CpG islands microarrays (CGI). Although each of them has its own advantages none of them is suited for large-scale screening as all are rather inefficient, complicated or technically challenging; they can be used only for testing a few samples. For example, after analysis of 1000 clones isolated using MBD/SPM, nine DNA fragments were identified as CpG islands and only one was specifically methylated in tumor DNA. Maximum resolution of RLGS is 1000–2000 NotI boundary clones.
Identification and quantification of microbial species in their various habitats is very important for understanding and dealing with different aspects of human and animal health and disease. For example, identification of pathogenic bacteria in food, soil or in air can prevent epidemics. Not very much is known about the human normal microflora. The human intestinal tract harbors a densely populated, active and complex bacterial ecosystem. The number of microbial cells in the colon is estimated to be 10–100 times larger than the number of eukaryotic cells in the human entire body and weighs more than one kg. Microscopic investigations demonstrated that many different species of microorganisms live in our digestive system, but at least 85% are unknown, mainly because they cannot easily be cultured in vitro. At the same time many studies have shown that the composition of the gut microflora plays a very significant role for human health. Many intestinal bacteria are known to provide molecules that the host itself cannot manufacture or degrade from nutritional compounds. Thus, these organisms are clearly of survival value, they are true symbionts of the host. At the same time, several major diseases are believed to have the gut flora as the main potential source of pathogenesis (allergy, IBD, Crohn disease, cancer, etc.). Identification of bacterial species and strains relies heavily on culture techniques. However, in a complex bacterial population, rapidly growing bacteria would overgrow, making quantification and identification of slow or non-growing bacteria impossible. Techniques still have to be developed even to culture a representative selection of the microorganisms. Consequently, the picture of the intestinal flora has been biased in favor of the more easily cultured bacteria. There are some methods available to analyze complex microbial mixtures, e.g. by enzyme analysis which requires growth of colonies outside the body, or analysis of the fatty acids composition in stools, both of which give crude and indirect indications of the composition of the normal flora. The limitations of such techniques are obvious. The application of culture-independent techniques based on molecular biology methods can overcome some shortcomings of conventional cultivation methods. In recent years the approach based on PCR amplification of 16S rRNA genes has become both popular and very useful. One modification of the approach utilized fingerprinting of all the species in the gut using, for instance, denaturing gradient gel electrophoresis (DGGE) with PCR amplified fragments of 16S rRNA genes. In another application, PCR amplified fragments of 16S rRNA genes were directly cloned and sequenced. These studies provided important information, however intrinsic disadvantages of the approach limit its application. The problem is that 16S rRNA genes are highly conserved and therefore the same sequenced fragment sometimes can represent different species. It is also difficult to adapt for quantification. Moreover, in fingerprinting experiments similar fragments may represent different species and yet different fragments may also represent the same species.

Microarray technology using immobilized DNA has opened up new possibilities in molecular biology of eukaryotes and prokaryotes. This approach was also applied to the studies of the bacterial composition of the microflora and identification of specific microbial species. However, the microarrays based on 16S rRNA genes suffer from the same problems as sequencing/fingerprinting methods and species-specific microarrays based on PCR amplification of specific DNA/gene fragments can only be used for identification of a limited number of microorganisms.

Methods

We suggested to use for the analysis of genomes and genome mixtures new methods: NotI passporting (tags) and NotI microarrays that were described in detail in (Li et al., 2002; Zabarovska et al., 2003; Zabarovsky et al., 2003). Important to note that for this analysis we use only some specific fragments of the genomes (NotI representations). Thus we do not aim to sequence all genomes or study all genes. We append special signatures for a particular organism/genes and analyze these signatures in different samples. Using the same idea (short sequence tags) we have also developed a new approach to genome mapping and sequencing based on slalom libraries.
The concept represents alternative approaches to the construction of linking and jumping libraries, and involves the construction of “slalom libraries”. The pilot experiments (Zabarovska et al., 2002) demonstrated the feasibility of the approach, and showed that the efficiency (cost-effectiveness and speed) of existing mapping/sequencing methods can be improved at least 5- to 10-fold. Furthermore, since the efficiency of contig assembly in the slalom approach is virtually independent of sequence read length, even short sequences, as produced by rapid high-throughput sequencing techniques suffice to complete a physical map and sequence scan of a small genome. Combination of these new sequencing techniques with slalom approach increase the power of the method 10–50 times more.

Results and Discussion

NotI microarrays for genome-wide screening cancer cells. The fundamental problems for genome-wide screening using NotI clones are: (i) the size and complexity of the human genome; (ii) the number of repeat sequences; and (iii) the comparatively small size of the inserts in NotI clones (on average 6–8 kb). To solve this problem, the special primers were designed and special procedure was developed to amplify only regions surrounding NotI sites, so called NotI representation (NR). Other DNA fragments were not amplified. We suggested to use for genome screening NotI microarrays in combination with this new method for labeling genomic DNA where only sequences surrounding NotI sites are labeled. A pilot experiment using NR probes demonstrated the power of the method, and NotI clones deleted in cancer lines, renal and breast biopsies were found. Experiments demonstrated that sensitivity of the method is enough to detect difference 1:2 (e.g. man from woman with the use of X chromosome specific NotI clones). Important to mention that in these experiments polymorphic and methylated NotI clones were also successfully detected. Our estimation is that human genome contains 10,000–15,000 NotI sites and 5,000–9,000 of them are unmethylated in a particular cell. Thus screening with NotI microarrays will be equivalent to screening using 3,000–4,000 gene associated single nucleotide polymorphisms (SNP). NotI microarrays give additional information to the deletion mapping: they can be used for gene expression profiling and methylation studies.

Moreover, NotI microarrays are the only existing microarrays giving the opportunity to detect methylation and copy number changes simultaneously or differentially. There is no reason why NotI microarrays cannot be used to study histone modifications and we are currently performing these experiments. NotI microarrays have another strong advantage compared to cDNA microarrays. There is no standard for comparing expression profiles. RNA is not a stable molecule and physiological conditions of the cancer cells and normal cells can be rather different, moreover such conditions vary very significantly during the short time period, depending on many different factors, e.g. temperature, day time, psychological status, medicine treatment, etc. On the other hand, with expression microarrays it is practically impossible to find the first events and first genetical lesions that leads to the development of cancer. This is not a problem for NotI microarrays as genetic lesions (for example deletions) are irreversible, epigenetic changes (e.g. methylation) are not so temporal and the normal genomic DNA is an perfect standard for comparisons. These features are rather important for different studies like diagnostics, prognosis, prediction etc.

NotI microarrays for analysis of complex microbial systems. In a pilot experiment we have produced NotI microarrays from gram-positive and gram-negative bacteria and have shown that even closely related E. coli strains can be easily discriminated using this technique. For example, two E. coli strains, K12 and R2, differ in less than 0.1 % in their 16S rRNA sequences and thus the 16S rRNA sequence would not easily discriminate between these strains. However, these strains showed distinctly different hybridization patterns with NotI microarrays. The same technique can be adapted to other restriction enzymes as well. This type of microarray opens the possibility not only for studies of the normal flora of the gut but also for any problem where quantitative and qualitative analysis of microbial (or large viral) genomes is needed.
**NotI passporting (generation of NotI tags) for analysis of complex microbial systems.** We demonstrated that these tags comprising 19 bp of sequence information could be successfully generated using DNA isolated from intestinal or faecal samples. NotI passports allow the discrimination between closely related bacterial species and even strains. This procedure for generating restriction site tagged sequences (RSTS) is called passporting and can be adapted to any other rare cutting restriction enzyme. A comparison of 1 312 tags from available sequenced *E. coli* genomes, generated with the NotI, PmeI and SbfI restriction enzymes, revealed only 219 tags that were not unique. None of these tags matched human or rodent sequences. Therefore the approach allows analysis of complex microbial mixtures such as in human gut and identification with high accuracy of a particular bacterial strain on a quantitative and qualitative basis. Among all tags from all sequenced bacteria 97% are species-specific.

**References**


