State-of-the-Art Lecture:
Genetics and genomics: implications for clinical practice

8

A. S. PÉÑA, M. L. LAINE, J. B. A. CRUSIUS, S. A. MORRÉ,
U. VAN DER VELDEN, F. LESSMANN, C. J. VAN DER PALEN,
A. J. VAN WINKELHOFF and B. G. LOOS

INTRODUCTION

Genetic predisposition contributes to the pathogenesis of most common diseases and several genes have been identified. Evidence for other susceptibility loci is also strong. Thus, the polygenic nature of inflammatory chronic diseases has been confirmed, but this does not follow Mendel’s rules. The genes so far identified in Crohn’s disease are involved in bacterial sensing such as CARD15[1–3]. Others are in the scaffolding of the intestinal epithelial cells, or in the transport of cations[1–3] or in the production of defensins[4]. These genes appear to be important for Caucasian patients only, and affect a subgroup of patients. This is difficult to understand and has delayed the implementation of genetic knowledge in clinical practice. Although substantial progress in the development of laboratory and analytical approaches to study non-Mendelian complex genetic disorders has been made, little is known of the interaction of several genes; first of genes which appear to function through complex networks and second the involvement of environmental factors and behavioural process that contribute to post-genomic regulatory events. These interactions are difficult to understand in spite of the accomplishments of the past few years.

The identification and characterization of the genetics and genomics of chronic inflammation are likely to contribute to understanding the basic aetiology of multifactorial diseases, such as inflammatory bowel diseases (IBD). This knowledge will improve risk assessment and influence therapeutics. Genetics identifies disease-related susceptibility genes while genomics identifies genes that belong to similar families based on their sequence homologies. The knowledge generated by genetics and genomics is now part of genomic science; this science is in the process of being integrated in
the clinical setting. Through the use of genetics, clinical scientists will be able to profile variations between individuals’ DNA to improve diagnostic skills, determine prognosis and predict responses to environmental and behavioural factors, as well as responses to particular drugs. This knowledge will contribute to a better understanding of common diseases which are multifactorial and polygenic (Table 1).

Transferring genomic science into a clinical setting is not easy. Technical and cultural challenges will have to be conquered to be able to use the new powerful genetic research tools. These tools are now available to classify the heterogeneity of disease and the individual responses to drugs. Also much of the generated genomic data of clinical relevance has been stored in a format that is inappropriate for use in routine diagnostic tests. The technology has not yet been established for rapid, inexpensive typing of most genomic biomarkers, with the exception of single-nucleotide polymorphisms (SNP). For routine clinical practice, current methods need to be simplified to effectively use the available advanced technology networks.

The first chip-based diagnostic and chip-based risk assessment tools for prognosis are now available: the chip-based diagnostic tool for predicting cytochrome P450 enzymes responsible for the oxidative metabolism of many drugs is a useful indicator of therapy response. The chip-based risk assessment tool, e.g. the recently introduced IBDchip, permits assessment of prognosis in subgroups of patients suffering from Crohn’s disease and ulcerative colitis. These tools will be implemented if there is clear evidence for their clinical utility. This technology has to be supported by both physicians and health authorities. Specific training in this field will need to be provided to health-care professionals in order to effectively use these new tools. It is precisely at this interface that biomedical informatics is of ultimate relevance. As recently expressed by Martin-Sanchez et al., a characteristic of the ‘post-genomic’ era will be to correlate essential genotypic information with expressed phenotypic information.

Medical informatics as we have known the discipline in research and development of informatics-based methods has produced tools for medical and epidemiological research, patient care and health management. Bioinformatics has grown up with the human genome project. Bioinformatics handles genomics, proteomics and any other biological research data. These two disciplines have resulted in biomedical informatics (BMI), an emerging discipline that aims to make possible the necessary translation of multiple systems with different databases into a common database. BMI will play an
important role in translating new insights in the pathophysiology of inflammatory processes and will help to develop new classification systems. It is also expected to provide information to design new treatment strategies\textsuperscript{6-8}.

To illustrate the state-of-the-art of the integration of genetics, genomics and BMI, in Amsterdam we are using as a model ‘periodontitis’, since this is also a common complex disease that requires the integration and analysis of multiple sources of data. Periodontitis offers a remarkable set of challenges to the new discipline BMI. The main challenge for BMI applied to complex diseases is the creation of a data warehouse (DW) for each disease to be studied. The INFOBIOMED consortium, a Network of Excellence of the European Union, has selected periodontitis as a pilot project (Figure 1). Periodontitis provides a good example of a chronic complex disease in which microbiological and genetic factors interact with environmental factors to determine disease susceptibility, severity of inflammation and evolution of the disease process.

PERIODONTITIS

Periodontitis is a chronic inflammatory disease of the supporting tissues of the teeth. These may show exposed root surfaces in swollen gums that easily bleed. The inflammatory process, a Th1 type of response, produces bone loss around the teeth which, after becoming loose, will eventually exfoliate. Patients with periodontitis experience problems with chewing due to tooth mobility and loss of teeth; they have bad breath and suffer from important subjective and objective aesthetic problems. Smoking is a risk factor strongly associated with periodontitis. In a study performed in Barcelona it was shown that smokers had 2.7 times, and former smokers 2.3 times, greater probabilities of having periodontitis than non-smokers. This was more evident after 10 years of smoking\textsuperscript{9}. Diet and exercise are important in maintaining healthy gums\textsuperscript{10-12} and stress enhances the risk of suffering from periodontitis (Figure 2).

The following oral bacterial species are recognized as periodontal pathogens: Porphyromonas gingivalis, Prevotella intermedia, Tannerella forsythensis, Fusobacterium nucleatum, Actinobacillus actinomycetemcomitans and the spirochaetal species Treponema denticola\textsuperscript{13}.

Genetic polymorphisms in a candidate gene approach have been explored as risk factors for periodontitis. There is limited evidence that some polymorphisms in the genes encoding interleukins (IL)-1, Fc gamma receptors (Fc gammaR), IL-10 and the vitamin D receptor, may be associated with periodontitis in certain ethnic groups, but the studies are underpowered\textsuperscript{14}.

These observations suggest that periodontitis may serve as a model for complex chronic inflammatory diseases, such as Crohn’s disease and ulcerative colitis. A periodontitis data warehouse (PDW) has been structured with the INFOBIOMED network and allows an integrative research approach, in which the new field of BMI will contribute to new insights in chronic inflammatory and infectious diseases (Figure 3).

The development of DW tools that will become available for clinicians and researchers opens the path to new technical advances in data mining and visualization. The BMI approach will in the future assist IBD clinical scientists and clinicians in research.
Figure 1  The EC-funded INFOBIOMED Study (Network of Excellence EC-IST 2001-35024)
Figure 2  Periodontitis is a multifactorial disease

Figure 3  The main challenge for biomedical informatics applied to periodontitis is the creation of a data warehouse
IBD – DIAGNOSTIC AND THERAPEUTIC STRATEGIES

BIOMEDICAL INFORMATICS

BMI is the emerging discipline that aims to put these two worlds together so that the discovery and creation of novel diagnostic and therapeutic methods is fostered\(^6\). In working with the model it appears that the challenges that BMI faces to be able to contribute to the advancement of the field can be summarized as follows (Figure 4):

![Diagram of BMI contributions and data flow]

**Figure 4** Development of data warehouse tools that will become available for clinicians and researchers opens the path to new technical advances in data mining and visualization.
GENETICS AND GENOMICS: IMPLICATIONS FOR CLINICAL PRACTICE

1. Data collection in daily clinical practice that responds to five distinct levels: genetic, environmental (smoking), infectious, disease phenotype.

2. Annotating and standardizing data of the different phenotypes determined by clinicians.

3. Annotating and standardizing data obtained by a dental image analysis (DIA) tool from dental X-rays. The radiographic data are kept in a particular protocol, ‘DICOM’, that can be accessed by any computer system in order to be able to analyse with multiple databases.

4. Ethical aspects such as ensuring privacy. The phenotype of the process, genetic data, data on the infectious component and environmental factors generated by dentists and dental researchers, as well as researchers in the field of informatics, should be encoded to ensure anonymization. This process should be reversible to be able to give advice to the patient after the studies have led to useful therapeutic actions. In the interactive (reversible) pseudonymization model from Custodix\footnote{15}, a transparent intermediary privacy protection engine is put between the users and the database web server (Figure 5).

5. Mining and visualization data analyse the different sets of data and then extract the meaning of the data. Data analysis and discovery learning algorithms produce a particular enumeration of models\footnote{16}. Data mining tools and knowledge discovery techniques are applied to the PDW.

The PDW has large quantities of heterogeneous data collected from diverse sources mentioned above, and this information is stored in specific categories so that it can be more easily retrieved, interpreted, and sorted by users (Figure 6).

![Diagram showing the privacy-enhanced data storage model of periodontitis ACTA (NL), Custodix (Belgium) and UPM (Spain)]

**Figure 5** Privacy-enhanced data storage model of periodontitis ACTA (NL), Custodix (Belgium) and UPM (Spain)
IBD – DIAGNOSTIC AND THERAPEUTIC STRATEGIES

Table 2  Main aims of data analysis, mining and visualization

| Clustering | New classification of periodontitis |
| Classification | New clinical and genetic associations |
| Association | New scientific hypotheses on pathophysiology of periodontitis |
| | New insights/pathways in other complex chronic inflammatory diseases |

INSIGHT AND EXPECTED RESULTS

To identify interesting patterns and model profiles, to discriminate between patients with periondontitis with different prognoses, and to identify similarities across patient records and groups, similar records to clusters programmes are being developed by the Institute of Computer Science (ICS) and Hellas (Forth) in Greece under the leadership of V Moustakis and G Potomias.

The main message that we would like to pass on is that BMI has the tools to integrate the different branches of knowledge, but patience and dedication are necessary to achieve results, and expertise from different areas is required. The main aims of the data analysis, mining and visualization are shown in Table 2.

APPLICATION OF THE MODEL TO IBD

The management of chronic inflammatory diseases, such as IBD, is a major challenge for health systems in any country. The aetiology is multifactorial and polygenic. A classification based on molecular biology, genetic information, phenotype information and modern imaging systems is bound to produce new ways to identify subgroups of patients with different prognosis. Genetics genomics and BMI need to be integrated into current research in IBD.

CONCLUSIONS

The INFOBIOMED model is expected to deliver improvements in better classification of periodontitis, and to suggest a new understanding of a multifactorial complex disease. It is to be hoped that the experience gained with this model will form the basis to encourage clinical research in inflammatory bowel diseases. Table 3 summarizes the contribution of the approach to these diseases.

Figure 6  Example of the heterogeneity of the data that is integrated in a data warehouse
IBD – DIAGNOSTIC AND THERAPEUTIC STRATEGIES

Table 3  Biomedical informatics will make possible the necessary translation of multiple systems with different databases into a common database

- Crohn's disease
- Ulcerative colitis
- Psoriasis
- Primary sclerosis cholangitis
- New comprehensive classification of inflammatory bowel disease
- New clinical–genetic associations
- New scientific hypotheses on pathophysiology of Crohn's disease and ulcerative colitis
- New insights/pathways in complex chronic inflammatory diseases

References


10