12.

ALU SEQUENCES IN THE HUMAN RESPIROME

David J Maguire, Harald Oey and Michael McCabe *

1. INTRODUCTION

The concept of the 'ome' has dominated some scientific literature in recent years since the introduction of the various 'genome' projects began in the late eighties. In a presentation at a recent conference on inborn errors of metabolism (ISIEM 2003), almost fifty such examples were cited. The concept is a useful one in so far as it serves to amalgamate and expedite access to a body of specific information within an area of scientific interest. The application of the concept is perhaps best illustrated by the published outcomes of the human genome projects in both the printed versions and by the easily accessible web-based version. The latter is particularly convenient in terms of ease of revision of information and also in terms of ability to link to related sites. It is the purpose of the current paper to highlight the need to develop and maintain a web-site that encompasses the information of interest to the members of the International Society on Oxygen Transport to Tissues (ISOTT). The structure of the 'respirome' and the specific components of this concept is presented in broad terms. As an example of the facility of such a resource, an example is given that encompasses the search for particular DNA sequence motifs (Alu sequences) within the human respirome. Analysis of the results of such searches can often provide insights into fundamental concepts. In this instance the search highlights the relative paucity of Alu repeats and an almost complete lack of inverted Alu repeats in exonic regions of some components of the human respirome elements investigated.

2. CASCADE MODEL OF OXYGEN FLOW FROM AIR TO CELL; A SCAFFOLD FOR THE HUMAN RESPIROME

In 1992, Erdmann presented a model that encompassed most of the information that represents the concept of the human respirome. He described the flow of oxygen from air to cell as a general model that was applicable to creatures as diverse in environments and mechanism of respiration as fish and humans. This model adequately describes the first part of the oxygen transport process. However, it fails to account for some of the other important aspects of the role of oxygen in sustaining life, namely the processes of oxygen transport and respiration within cells and the subsequent steps of synthesis and

* Genomics Research Centre, Griffith University, Nathan, Brisbane, Australia, 4111

detoxification of injurious oxygen metabolites. It is also considered important to include the processes by which inadequate control of those toxic components lead to cell death. In this context, processes that protect cells against those toxic products are also relevant. A number of molecules that impact upon the processes described above have only been recognized since the human genome publications at the end of the last millennium and rate special mention. A modified version of Erdmann's model[1] is presented elsewhere in this volume[2]. The modules that compose this expanded model are; oxygen transport within blood, oxygen exchange between blood and cells, oxygen transport within cells, cellular respiration, antioxidant processes, free radical generation and programmed cell death (apoptosis).

3. APPLICATION OF THE HUMAN RESPIROME; SEARCH FOR ALU REPEAT SEQUENCES

In order to illustrate the potential utility of the respirome, an investigation is reported into the Alu content of a subset of the genes encoding members of the respirome. Alu sequences are a sub-group of the Lines (long interspersed nuclear elements) that are almost exclusive to humans and their very close ancestral relatives. The ability of direct inverted Alu repeat sequences to form stable secondary RNA structures is believed to be an important step in the process of RNA editing by enzymes such as adenosine deaminases acting against RNA (ADARs)[3]. The action of these enzymes against pre-mRNA leads to the conversion of adenosine (A) to inosine (I). A recent genome-wide investigation into A to I editing sites has increased by two orders of magnitude the known incidence of such sites [4].

4. METHODS

A bioinformatics approach was used to search for Alu sequences and Alu repeat sequences in the set of genes specified in Table 1. The gene-coding sequences were obtained from the NCBI web-site online database of reference sequences (Refseq)[5]. When splice variants with differing 5' or 3' untranslated regions (UTRs) were available for a gene, the longest form of the gene was used. The mRNAs for the genes were aligned with the corresponding genomic DNA sequences obtained from NCBI (build 35.1). The sequences were analysed for presence of Alu repeats only, using the repeatmask program[6], specified to only mask Alus, and data were extracted from the output of these analyses. Alu fraction was calculated as the percentage of total length of Alu repeats in a gene relative to total gene length. After Alu repeat sequences were highlighted by this technique they were assigned to either intronic or exonic regions of the genes analysed. Further analysis of these results enabled predictions to be made regarding susceptibility to the editing enzymes known as ADARs. Total Alu length and total gene length in all the components of the human respirome examined in this study was calculated and expressed as a percentage.

5. RESULTS

The results of the bioinformatics approach to examination of the respirome are shown in Table 1 which shows that there is huge variability in gene length, mRNA length, exon content, Alu content and total Alu length. Some members of the respirome gene assemblage contain no such sequences, significantly the two human globin gene