

**Bioc471a/571a Homework 10 - Due at the start of class November 20****Lab Practicum - Mine Your Own Business****Background**

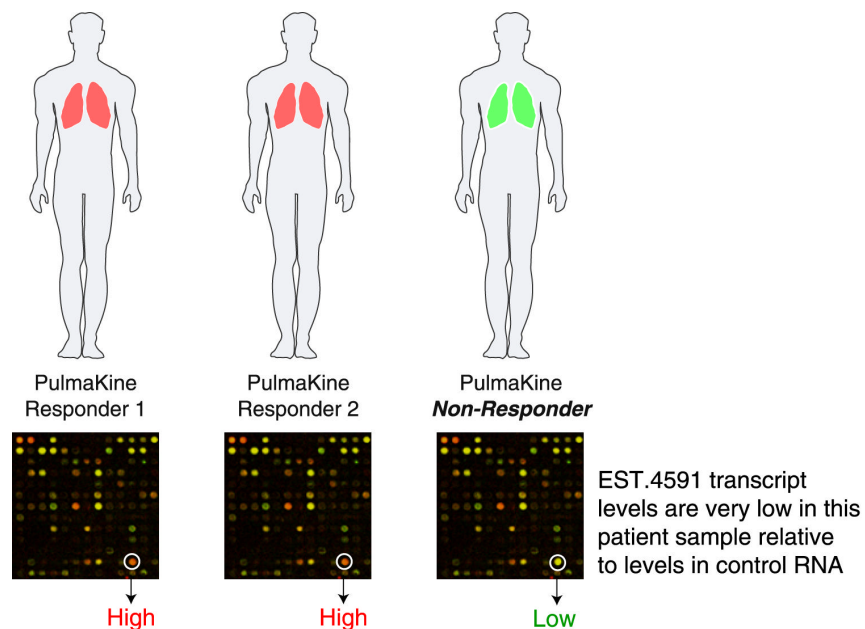
A pharmacogenetic-based biotechnology company, Ocard Asthma Inc., has developed a new treatment for acute asthma attacks it calls PulmaKine. This inhaler-delivered compound was discovered using a combinatorial chemistry approach directed against known cytokine receptors expressed on lung epithelial cells. During the clinical trial phase, it was observed that PulmaKine was highly effective in blocking respiratory impairment in 65% of asthmatics who used the Ocard inhaler during an acute asthma attack. However, the other 35% of the test subjects showed no measurable improvement with the drug, and in fact, many of these PulmaKine non-responders required emergency medical treatment to stop the asthma attack.

Bronchial lavage fluid (BLF) had been collected from all of the study subjects. RT-PCR microchip RNA expression profiling was performed using the BLF cell samples. This characterization led to the identification of a cDNA transcript (EST.4591) that was found to be expressed in the lung tissues of PulmaKine Responders, but not expressed in the PulmaKine Non-responders. Genotyping of the EST.4591 genomic locus identified a germline polymorphism in PulmaKine Non-responders that explained the lack of expression in BLF samples. The researchers named the putative EST.4591 gene product AMG (asthma modulating gene).

**Basic Strategy**

The molecular genetic research group at Ocard intend to use the microchip data as a starting point to investigate the function of the AMG protein, and moreover, to develop a diagnostic test that can be used to positively identify asthma patients who will benefit from PulmaKine therapy. One long term goal of the company is to develop a second generation therapeutic compound that would be used to specifically treat PulmaKine Non-responders that exhibit low levels of AMG expression.

The illustration below shows how data will be collected from PulmaKine patients based on RNA Expression Profiling using human cDNA microarrays. In this image analysis, red spots indicate that the specific transcript is expressed at a higher level in the patient sample than it is in the control RNA, whereas, green spots denote lower expression of a particular gene in the patient sample relative to the control.



BLAST searches were done with EST.4591 to try and establish a functional identity. Initial results suggested that EST.4591 may corresponded to non-coding sequence based on low homology (E values >0.001) to sequences in the "nr" Database.

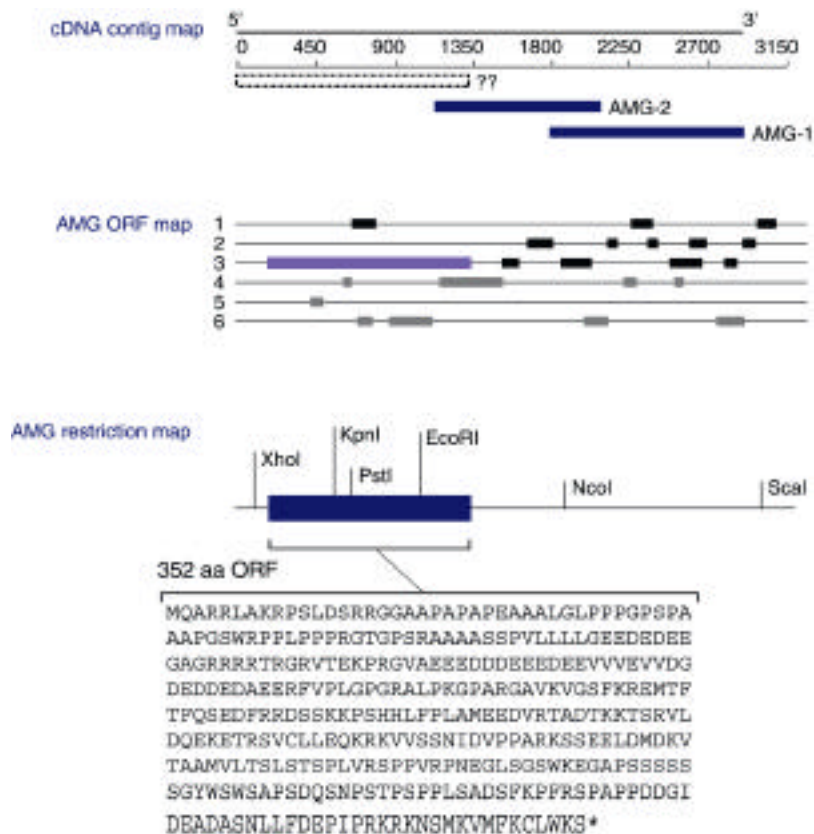
**EST.4591**

5' - TTCCTGAGAACAGTGGGAGCCTTGGGCAGGTGOGGAGAATCTGACCTGCTGGATTTCATGACGTGG  
 ATTTCTTGAGCTGCAGCTGGGCGGTAGGCCCGGGGCCCGCGGACGTCCAGTACGACCTGTACTTGAA  
 CGTTGCCAACAGGCCTCAACAGTACGAGTGTCTTCACTACAAAACGGATGCTCAGGGAACACGTATCGGG  
 TGTGCTTTCGATGACATCTCTCGACTCTCCAGCGGTTCTCAAAGTTCCACATCCTGGTGCGGGGCAGGA  
 GCGCAGCCTTCGGTATCCCCTGCACAGATAAGTTTGTGCTCTTTTCACAGATTGAGATATTAATCCACC  
 CAACATGACTGCAAGTGTATAAGACACATTC - 3

BLAST search results using EST.4591 with *blastn*

Sequences producing significant alignments:	Score (bits)	E Value
<a href="#">ref NM_016973.1 </a> Mus musculus sialyltransferase 7 ((alpha-N...	40	0.21
<a href="#">emb Z93241.1 HS22E13</a> Human DNA sequence from clone RFI-22...	40	0.21
<a href="#">db AB035123.1 AB035123</a> Mus musculus ST6GalNAc VI mRNA for ...	40	0.21
<a href="#">db AB035174.1 AB035174</a> Mus musculus ST6GalNAc VI mRNA for ...	40	0.21
<a href="#">gb AC084554.1 CBRG34116</a> Caenorhabditis briggsae cosmid G34L...	38	0.85
<a href="#">gb AE003491.2 AE003491</a> Drosophila melanogaster genomic scaff...	38	0.85
<a href="#">emb X12426.1 X12426</a> Xenopus laevis UI 70K gene exon 4, 5,...	38	0.85
<a href="#">gb U28738.1 CELT28D9</a> Caenorhabditis elegans cosmid T28D9	38	0.85
<a href="#">gb AC018720.5 AC018720</a> Homo sapiens clone CTD-2562F8, compl...	36	3.4
<a href="#">gb AC008276.4 AC008276</a> Homo sapiens clone RP11-29713, compl...	36	3.4

Sequence homology searches with the AMG-1 sequence (EST.4591) were performed with other GenBank databases to identify overlapping EST sequences that could be used to extend the AMG transcript sequence. The largest human cDNA identified by these "database mining" approaches was named AMG-2 and found to extend within ~1200 base pairs of the predicted 5' end based on Northern blots. No other GenBank sequences were found to extend beyond AMG-2. In order to isolate the missing portion of AMG, the researchers mail-ordered the AMG-2 cDNA from the public domain storehouse (I.M.A.G.E. Consortium) and used it as a reagent in the lab to isolate a third clone called AMG-3. DNA sequence analysis of the combined AMG sequence (AMG contig) identified a single AMG open reading frame of 352 amino acids.



1. (3 pts.) How are the fluorescently labeled Cy3 (control) and Cy5 (experimental) nucleic acids generated for use in expression profiling studies by microarray technology? What type of RNA sample might have been used to generate the Cy3-labeled control nucleic acid in order to produce the observed microarray results?

**The nucleic acid is cDNA and it is labeled by incorporating Cy3 or Cy5 labeled dNTPs in an in vitro reaction using reverse transcriptase and template RNA from the patient or control samples.**

**The purpose of the Cy3-labeled "control" RNA is to serve as a standard against which all patient RNA samples can be compared against. This type of cell-specific control could have been derived from any readily available cell or tissue source, as long as it expressed low levels of AMG transcript. For example, a human lung cell line or a collection of BLF RNA samples from non-asthmatics. The level of AMG RNA must have been higher in the control RNA than in the PulmaKine Non-Responders to give the observed results.**

2. (2 pts.) The GenBank "blastn" search of EST.4591 against the "nr" database did not show any significant matches. Based on the available NCBI databases, list and briefly describe two other GenBank databases that could have led to the identification of AMG-2 as an overlapping, but uncharacterized, AMG expressed sequence tag (EST).

***dbest* - non-redundant database from EST divisions.**

***human ests* - non-redundant EST database limited to human sequences.**

***human genome* - all database sequences released to the human genome.**

3. (2 pts.) Describe two independent methods that the Ocard researchers could have used in their molecular genetic lab to isolate the AMG-3 cDNA sequence.

**1) Screen a human lung cell cDNA library using the 5' end of the AMG-2 clone as a probe to find overlapping clones.**

**2) Use RTase and 5' RACE to extend RNA isolated from BLF samples using primers derived from AMG-2 sequence. Clone the resulting PCR products by TA cloning or by restriction enzyme sites included in the PCR primers. The AMG-2 clone was used as a reagent in the lab to set up a positive control for the 5' RACE strategy.**

4. (3 pts.) Describe a GenBank query that could be used to identify a possible function for the 352 amino acid AMG protein? Assuming you get a significant match to a related gene from a model organism, what other types of databases could you query through the Internet to find out more about the function of AMG?

**Use the AMG protein sequence in a "tblastn" GenBank search against the nr database to identify protein homologies. One could also use a Psi blast search which utilizes a different algorithm for protein searches.**

**Assuming an AMG ortholog was found in yeast, Drosophila or C. elegans, a query of the integrated genome/genetic/cell biology database for that organism could be done to see if there is a genetic mutation and phenotype previously described for that gene. Follow-up studies using the relevant PubMed articles would provide information about laboratories that might be willing to function as collaborators in new studies aimed at developing second generation asthma therapies.**