Spring 2011 Genomics Exam #2
Sequence Variations and Microarrays

There is no time limit on this test, though I don't want you to spend too much time on it. I have tried to design an exam that will take about the same amount of time as the first one. You are not allowed to read any papers to help with this exam. There are 6 pages, including this cover sheet, for this test. There are no Discovery Questions on this exam. You are not allowed discuss the test with anyone until all exams are turned in at 9:30 am on Wednesday March 23. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE AT 9:30 am ON WEDNESDAY MARCH 23**. You may use a calculator, a ruler, your notes, the book, and the internet. You may take this exam in as many blocks of time as you want. Submit your electronic version before 9:30 am (eastern time zone).

The **answers to the questions must be typed in a Word file and emailed to me as an attachment**. Be sure to backup your test answers just in case (I suggest a thumb drive or other removable medium). You will need to capture screen images as a part of your answers which you may do without seeking permission since your test answers will not be in the public domain. Remember to explain your thoughts in your own words and use screen shots to support your answers. Screen shots without ***your*** words are worth very few points.

DO NOT READ or DOWNLOAD ANY PAPERS FOR THIS EXAM. RELY ON YOUR EXPERIENCE, AND YOUR SKILLS.

-3 pts if you do not follow this direction.
Please do not write or type your name on any page other than this cover page.
Staple all your pages (INCLUDING THE TEST PAGES) together when finished with the exam.

Name (please type):

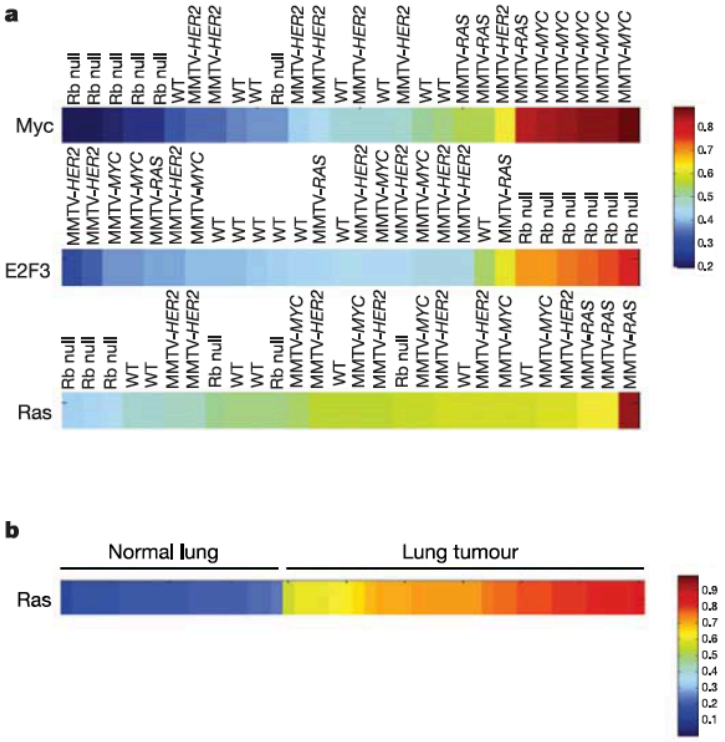
Write out the full pledge and sign (electronic signature is ideal):

How long did this exam take you to complete?

blue red

16 points (4 pts each)

1) Investigators engineered human cells to over-express three oncogenes (Ras, Her2 or Myc) or they deleted the tumor suppressor Rb gene (both alleles). In panel a) they used Affy DNA microarrays to measure the activity of three different oncogenic pathways (Myc, E2F3 or Ras) and the color-coded each engineered cell line based on the probability of the 3 distinct pathways to be activated. In panel b) non-engineered cells isolated from human patients were analyzed as in a).



a) What can you conclude from the data in part b? Limit your answer to a 3 sentence maximum.

Lung cancers (of this type) appear to utilize the Ras pathway more than normal lung cells do. The range of investigator confidence varies from very high probability (0.9) to about 0.5 probability for the cancer cells. Normal cells have very low probability of Ras activation (≤ 0.2).

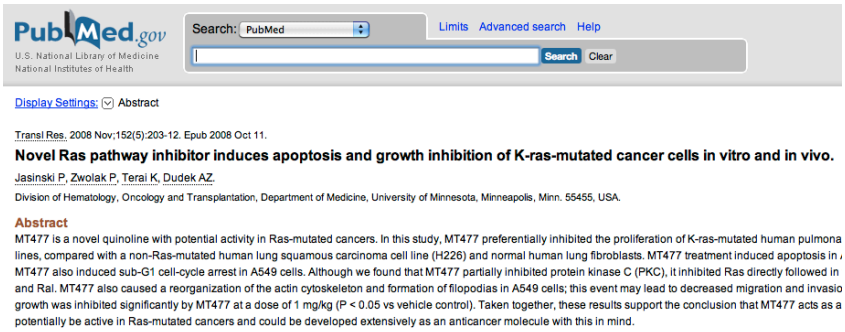
b) How could you use these data to personalize cancer treatment for any cancer type? 3 sentence maximum.

First, find drugs that kill each of these transformed cancers. Then match the patient biopsy to the type of engineered cell. Then use the drug that kills the engineered cells on the patients.

c) Use NCBI to find a new drug (not FTS) that might work especially well on cancers with activated Ras pathways. What drug did you identify? You must include a screenshot of your findings and the URL of the NCBI resource that you used to reach this answer.

MT477

http://www.ncbi.nlm.nih.gov/pubmed/19010291


PubMed.gov
U.S. National Library of Medicine
National Institutes of Health

Search: PubMed Limits Advanced search Help

Search Clear

Display Settings: Abstract

Transl Res. 2008 Nov;152(5):203-12. Epub 2008 Oct 11.

Novel Ras pathway inhibitor induces apoptosis and growth inhibition of K-ras-mutated cancer cells in vitro and in vivo.

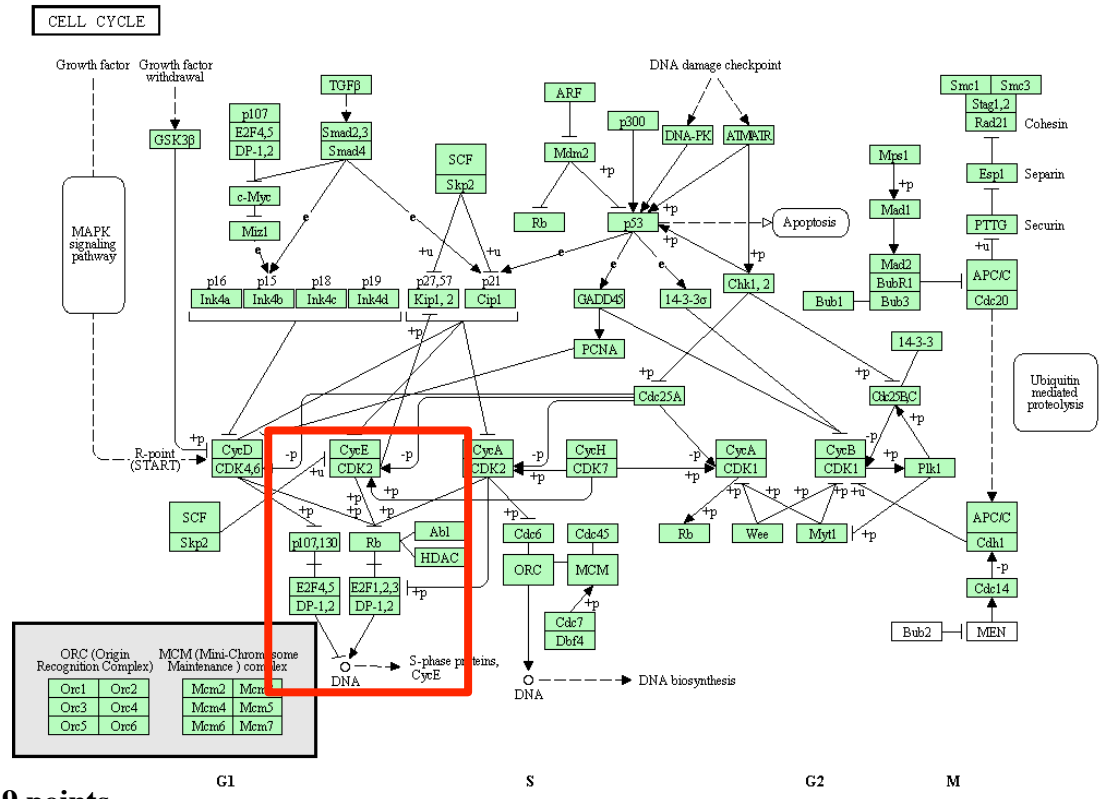
Jasinski P, Zwolak P, Terai K, Dudek AZ.

Division of Hematology, Oncology and Transplantation, Department of Medicine, University of Minnesota, Minneapolis, Minn. 55455, USA.

Abstract

MT477 is a novel quinoline with potential activity in Ras-mutated cancers. In this study, MT477 preferentially inhibited the proliferation of K-ras-mutated human pulmona lines, compared with a non-Ras-mutated human lung squamous carcinoma cell line (H226) and normal human lung fibroblasts. MT477 treatment induced apoptosis in MT477 also induced sub-G1 cell-cycle arrest in A549 cells. Although we found that MT477 partially inhibited protein kinase C (PKC), it inhibited Ras directly followed in and Ral. MT477 also caused a reorganization of the actin cytoskeleton and formation of filopodias in A549 cells; this event may lead to decreased migration and invasio growth was inhibited significantly by MT477 at a dose of 1 mg/kg ($P < 0.05$ vs vehicle control). Taken together, these results support the conclusion that MT477 acts as a potentially be active in Ras-mutated cancers and could be developed extensively as an anticancer molecule with this in mind.

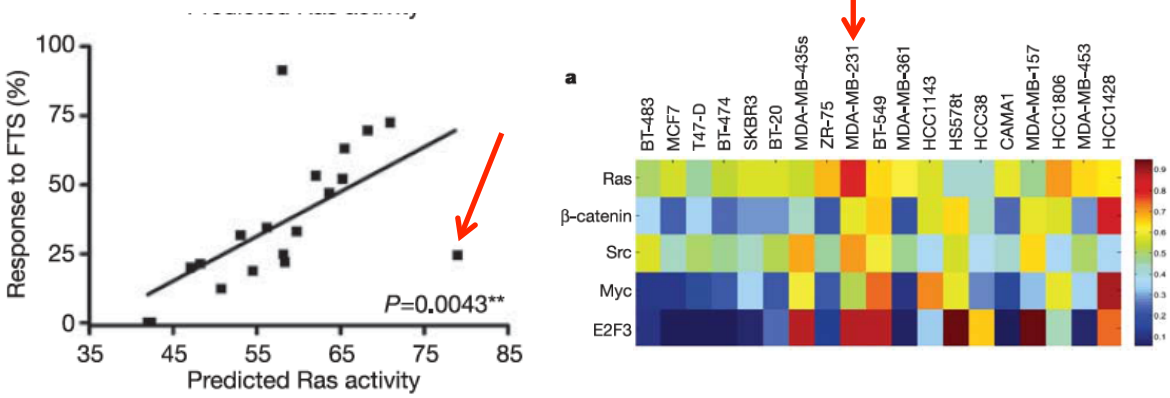
d) Find a visual pathway map that shows the relationship between Rb and E2F3. In 3 sentences or less, explain the data in the figure used in this question given what you found for your pathway map. Be sure to include a screen shot and the URL for the pathway map you find.
http://www.genome.jp/kegg-bin/show_pathway?hsa05223
It seems that when Rb was probably activated, E2F3 was not, and vice versa. The image below shows that these two proteins are antagonists, so in cancers, either one is activated or the other one is. They aren't both activated or they'd cancel each other out.



9 points

2) Another figure from the same paper looking at oncogene pathways and how to cure cancers.

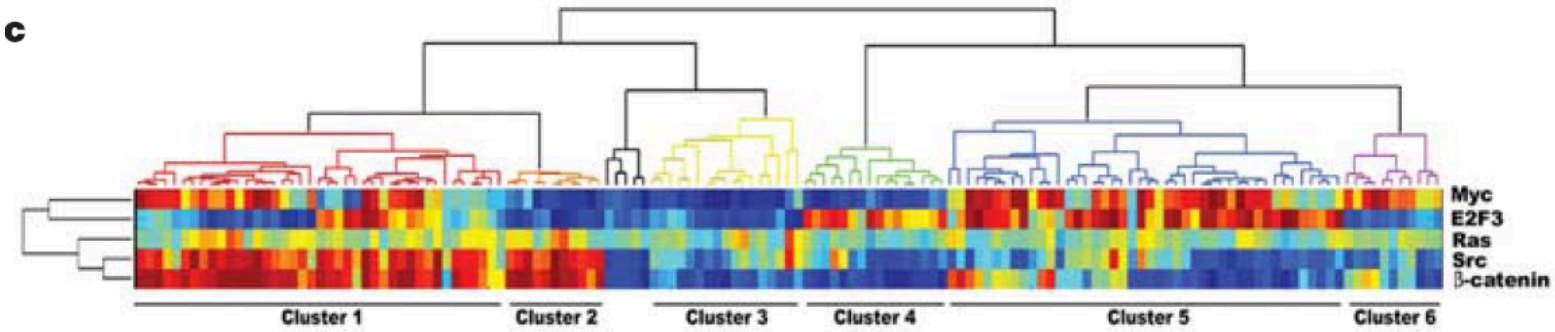
a) Use the data from the figures below to determine which sample in the multi-color figure is the one pointed to by the red arrow. Just name the sample.

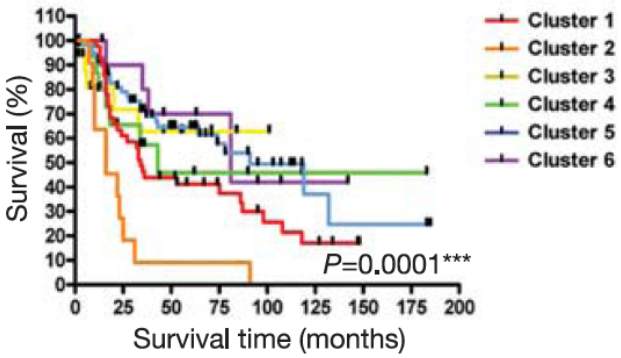


b) What conclusion(s) can you draw about personalized medicine based on the figures above? Limit your answer to 3 sentences or less.
We will be able to match pathway activity with appropriate drug treatment, much of the time but not always. Note the two outliers. The one high up was overly sensitive to the drug while the one in the bottom right corner was insensitive to it. Cancer is more complicated than just one pathway, but the idea of treating the right pathway is better than treating the cancer phenotype ignorant of the activated pathway.

c)) What can you deduce given the outliers in the black and white graph on the left below? Limit your answer to 3 sentences or less.
See above.

12 points (3 pts each)
3) Look the third and final figures (below) relating DNA microarrays to cancer treatment. As before, the heat map was used to indicate the probability of pathway activation. In this case, they have taken breast cancer biopsies and run Affy chips on them to determine which pathways might be activated.
a) How many types of breast cancer are depicted in this figure? In two sentences or less, explain how you reached your number.
7 types, 6 clusters plus the odd block as long as that is cancerous and not wt.
b) In two sentences or less, describe how the investigators cut the tree to generate their clusters.
I looks like they used the correlation coefficient for cluster 3 as the cutoff for all trees.
c) Use a numbered list to tell me the molecular pathway(s) in the most aggressive breast cancer subtype.
Src and β-catenin are the two activated pathways in the most aggressive cancer (cluster 2).
d) Which breast cancer subtype is the least lethal? Describe the 5 molecular pathways for this type of cancer in 3 sentences or less.
Cluster 3 is the least lethal and it looks like none of the pathways have high probability of activation, though the highest is Ras. The one exception is one of the two samples on the far right that have Ras and Src activated with high probability.





24 points (5 pts each except last one is 4 pts)

4) For this question, I am providing you with the abstract and one figure. The "questions" ask you to draw data that match the text. My goal is to measure how well you can understand the jargon-rich text and communicate the meaning in a non-verbal way. You may do this by hand and scan your images, or draw them electronically. I recommend you do a rough draft first. I need to be able to read your hand writing, so write neatly.

a) Draw a picture of well controlled experimental results showing the transcription levels of Cdkn2a and Cdkn2b in cardiac tissue.

wt – level

mutant – reduced

control with actin

b) Diagram what the abstract is describing about the experimentally produced heterozygous mice.

cis-acting mechanism

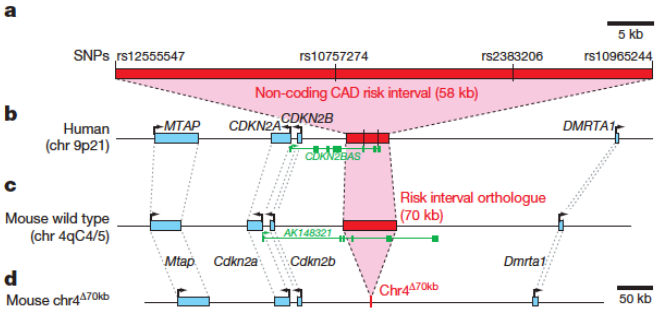
c) Diagram the experimental results showing what happens to aortic smooth muscle cells in homozygous deletion mice.

they proliferate – needs to be experimental results

d) Draw a picture of the cardiac blood vessels in a homozygous mutant mice that explains why CAD results from the engineered deletion.

occluded

e) In words, speculate why this deletion causes the phenotype it does. Limit your answer to 3 sentences or less. Enhancer, promoter in the intron of this green gene.

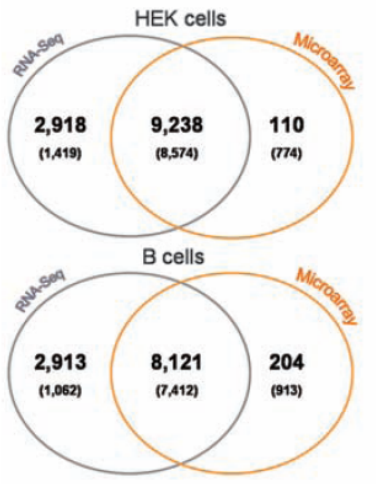


Sequence polymorphisms in a 58-kilobase (kb) interval on chromosome 9p21 confer a markedly increased risk of coronary artery disease (CAD), the leading cause of death worldwide[1,2]. The variants have a substantial effect on the epidemiology of CAD and other life-threatening vascular conditions because nearly one-quarter of Caucasians are homozygous for risk alleles. However, the risk interval is devoid of protein-coding genes and the mechanism linking the region to CAD risk has remained enigmatic. Here we show that deletion of the orthologous 70-kb non-coding interval on mouse chromosome 4 affects cardiac expression of neighbouring genes, as well as proliferation properties of vascular cells. Chr4$^{\Delta 70kb/\Delta 70kb}$ mice are viable, but show increased mortality both during development and as adults. Cardiac expression of two genes near the non-coding interval, *Cdkn2a* and *Cdkn2b*, is severely reduced in chr4$^{\Delta 70kb/\Delta 70kb}$ mice, indicating that distant-acting gene regulatory functions are located in the non-coding CAD risk interval. Allele-specific expression of *Cdkn2b* transcripts in heterozygous mice showed that the deletion affects expression through a *cis*-acting mechanism. Primary cultures of chr4$^{\Delta 70kb/\Delta 70kb}$ aortic smooth muscle cells exhibited excessive proliferation and diminished senescence, a cellular phenotype consistent with accelerated CAD pathogenesis. Taken together, our results provide direct evidence that the CAD risk interval has a pivotal role in regulation of cardiac *Cdkn2a/b* expression, and suggest that this region affects CAD progression by altering the dynamics of vascular cell proliferation.

12 points

5) This question is also a learning moment.

To the right are two Venn diagrams showing the number of genes detected by either direct RNA sequencing (RNA-seq) or by DNA microarrays. HEK = human embryonic kidney cell line. Bigger, bold numbers indicate the number of genes measured by the appropriate method at least once while the smaller (numbers) indicate genes measured at least five times.



a) What trend do you notice when comparing the two methods?

RNA seq detects more rare mRNAs than microarray.
Microarray detects more replicates

b) As a skeptical scientist comparing a new method with a more familiar method, what additional data would you like to see before determining if the new method was valid or not.

qPCR to validate and replicates +/- variation. (2 pts each)

c) What aspect of the data visualization is less than optimal if you are skeptical of this new method?

Presentation is odd too with size font.
wording is odd of at least 1 time appears to not include 5 times

16 points

6)

rs=2298771: **6 pts**

a) tell me what gene this is, what sort of genomic event is being described, what phenotype is associated with this number, and the population distribution in 4 distinct groups. To receive credit, you must provide me with all links to the sites you used. Remember, do not read any papers for this – use only credible databases.

http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=2298771

Voltage-gated sodium channel

http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene&cmd=Retrieve&dopt=Graphics&list_uids=6323

SNP G → A base, missense mutation from Ala → Thr.

Associated with epilepsy

Population Diversity

Sample Ascertainment					Genotype Detail NEW					Alleles	
ss#	Population	Individual Group	Chrom. Sample Cnt.	Source	A	A/A	A/G	G/G	HWP	A	G
ss117956421	YRI		2	IG			1.000			0.500	0.500
ss123876129	pilot.1.CEU		72	AF						0.625	0.375
	pilot.1.CHB+JPT		88	AF						0.875	0.125
	pilot.1.YRI		50	AF						0.840	0.160
ss138571186	ENSEMBL_Watson		2	IG		1.000				1.000	
ss164736813	YRI	Sub-Saharan African	2	IG			1.000			0.500	0.500
ss165561144	CEU	European	2	IG			1.000			0.500	0.500
ss167382317	PGP		2	IG		1.000				1.000	
ss201323589	BUSHMAN_POP		3	IG	0.500		0.500			0.667	0.333
	BUSHMAN_POP2		1	IG	1.000					1.000	
	BANTU		1	IG	1.000					1.000	
ss219633291	pilot_1_YRI_low_coverage_panel		118	AF						0.864	0.136
ss231453038	pilot_1_CEU_low_coverage_panel		120	AF						0.600	0.400
ss238944105	pilot_1_CHB+JPT_low_coverage_panel		120	AF						0.858	0.142
ss24299377	AFD_EUR_PANEL	European	48	IG		0.208	0.625	0.167	0.251	0.521	0.479
	AFD_AFR_PANEL	African American	44	IG		0.773	0.182	0.045	0.294	0.864	0.136
	AFD_CHN_PANEL	Asian	48	IG		0.750	0.208	0.042	0.439	0.854	0.146
ss48405418	AGI_ASP_population	multiple	66	IG		0.636	0.242	0.121	0.100	0.758	0.242
ss68834082	HapMap-CEU	European	120	IG		0.417	0.450	0.133	1.000	0.642	0.358
	HapMap-HCB	Asian	90	IG		0.756	0.244		0.371	0.878	0.122
	HapMap-JPT	Asian	90	IG		0.733	0.222	0.044	0.317	0.844	0.156
	HapMap-YRI	Sub-Saharan African	120	IG		0.767	0.217	0.017	1.000	0.875	0.125

gq343003 **2 pts**

b) Tell me what gene this is, what phenotype it is associated with, where the DNA came from.

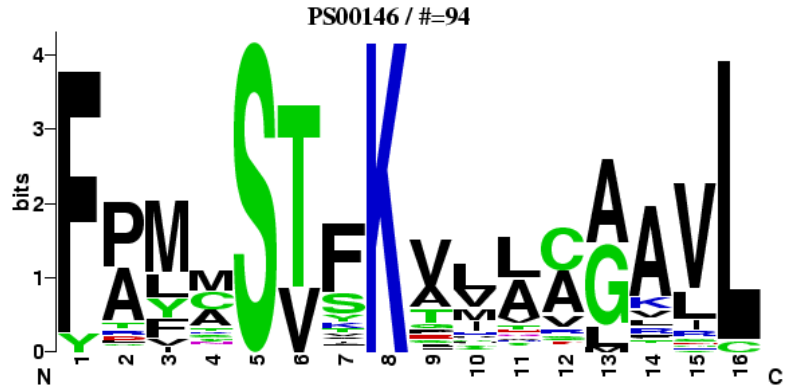
http://www.ncbi.nlm.nih.gov/nuccore/gq343003

beta lactamase gene (antibiotic resistance) from an uncultured microbe.

c) Find a screen shot of the sequence logo for this class of protein and determine how well gq343003 matches this sequence logo. **4 pts**

http://www.expasy.org/cgi-bin/prosite/sequence_logo.cgi?ac=PS00146

http://www.ncbi.nlm.nih.gov/protein/254966766?report=fasta



MKLWTSLCLFLLLGATAFAVRENPLESRLRRIIRLSNTDVGVAVISGSRNWAVGNRKR**PL**L**SVFK**FFI**A**V
QT**L**RQMEQTGTALNAKLTVEENMTDANMYSPMLKKHPRRPFEISLAELLEYMIAESDNNAADILLKYSGG
TGQLEKFLHDLGFSGIDIRVNEKQMNQKAENQYLNQAAPSDVAGLIKLVLEKDVLSAEHRKFLSDIMLKT
STGTDKIKQGLPPGVMFGHKTGSSSRTADGIKIADNDAGFVTLTNGRTYYIAVMITESKLDDRANAALAA
QISQTVYNYLTTEKTD

The motif is partially conserved. The red letters indicate amino acids that are visibly represented in the logo. Note that the first F is not present in this version.

d) What degree of amino acid identity does this protein have with its highest BLAST hit for a known species? Show me a screen shot of your best BLAST results of a known species.
4pts

```
>ref|ZP_03016044.1|  hypothetical protein BACINT_03645 [Bacteroides intestinalis DSM
17393]
 gb|EDV04508.1|  hypothetical protein BACINT_03645 [Bacteroides intestinalis DSM
17393]
Length=301

 Score =  214 bits (545),  Expect = 1e-53, Method: Compositional matrix adjust.
 Identities = 112/271 (41%), Positives = 168/271 (62%), Gaps = 3/271 (1%)

Query  25   LESRLRRIIRLSNTDVGVAVISGSRNWAVGNRK--RPLLSVFKFFIAVQTLRQMEQTGTA  82
            +E ++  +++     VGVAV++     AV N +   PLLS+FKF + +  L +M++   A
Sbjct  30   IEQQIDSLLKDKKATVGVAVLANDETVAVYNNQIHFPLLSIFKFHVGLAVLDKMDKGHIA  89

Query  83   LNAKLTVEENMTDANMYSPMLKKHPRRPFEISLAELLEYMIAESDNNAADILLKYSGGTG  142
            L++ + V+ +    N YSP+  K P +   ISL ELL+Y I++SDNN  DIL++Y GG
Sbjct  90   LDSLIEVKSSQLTPNTYSPLRDKFPDQNITISLGELLKYTISKSDNNTCDILIEYVGGIE  149

Query  143  QLEKFLHDLGFSGIDIRVNEKQMNQKAENQYLNQAAPSDVAGLIKLVLEKDVLSAEHRKF  202
            Q+ +++  LG    ++   E  M+   +  YLN + P +V  L+ +  ++ + +   +++ F
Sbjct  150  QVNEYVKSLGIKDCNLAATETLMHTSGD-AYLNWSTPEEVVRLLNIADKQPLFGTQYKDF  208

Query  203  LSDIMLKTSTGTDKIKQGLPPGVMFGHKTGSSSRTADGIKIADNDAGFVTLTNGRTYYIA  262
            L  IM +TSTG DK+K  LP  V+ GHKTGSS RT +GIKIADNDAGFV L NG+ YYIA
Sbjct  209  LQAIMQETSTGKDKLKGQLPADVIVGHKTGSSDRTPEGIKIADNDAGFVILPNGQKYYIA  268

Query  263  VMITESKLDDRANAALAAQISQTVYNYLTTE  293
            V + ES+  D  NAA+ A ISQ VY+ L ++
Sbjct  269  VFVMESQETDADNAAIIASISQIVYDTLNSD  299
```

11 points

7) This last question is pretty straight forward. Coincidentally, it is related to the bioinformatics talk this week!

a) What type of phylogenetic tree is shown below? What name is associated with this type of tree? (Same question worded two ways.) **3 pts**

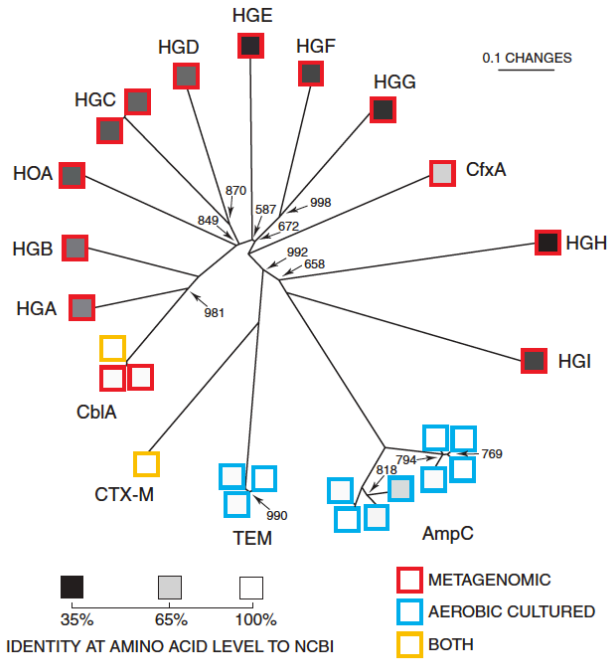
unrooted

b) To what family of molecules do AmpC, TEM, CTX-M and CblA belong? How did your reach this conclusion? **4 pts**

amp resistance

c) What is the medical consequence of this metagenomic study given that the HG_ series represent different examples from the same protein family? The DNA for the HG_ series was isolated from fecal samples taken from only two people. **4 pts**

There is more antibiotic resistance in humans than we had known about which means antibiotic resistance is a serious problem that we will not defeat any time soon. The amount of amino acid diversity in HG_ proteins is shockingly diverse. Engineering new antibiotics by tweaking is probably not going to work.



Bonus Question (3 pts):

What innovation did the group invent that allowed them to test every exon in the human genome for expression using DNA microarrays? inkjet printers