SISSA PhD Course in Genomics: admission exam, written test

Sept 13th, 2018

PRIOR TO INITIATE YOUR TEST, PLEASE READ CAREFULLY THE FOLLOWING POINTS:

(a) Pick two (*and not more*) questions among the following ones and answer them

(b) Use one (*or more*) sheets per answer. Label each sheet with the corresponding question number: "Question N". DO NOT use the same sheet for different answers.

(c) To allow for fair and unbiased evaluation of your elaborates, DO NOT label them with your name and surname.

(d) You can answer questions in Italian or English. Your language choice WILL NOT affect the evaluation of your test at all. If you choose Italian and pass the written exam, you will be asked to document your knowledge of English, by translating a short paragraph from Italian to English upon your oral exam.

(e) You can write a rough draft and then make a fair copy of it. Alternatively, if you prefer, you can straightly write a "fair copy". Whatever your choice is, pay special attention to clarity and intellegibility of your handwriting. All sentences of your "fair copy" not fullfilling the requirement of clarity and intellegibility WILL BE IGNORED AND NOT EVALUATED. In this respect, you are ENCOURAGED to write in block letters, even in CAPITOLS.

(f) You have three hours for your exam

(g) As you complete your elaborates, please put into the big envelope:
- the elaborates
- the small envelope, containing the stripe of paper with your name and surname
Both big and small envelopes MUST NOT report your name and surname at all.

PAY ATTENTION: FAILURE TO FULLY COMPLY WITH ISSUES (a), (b), (c), (d), (f) and (g) WILL CAUSE YOUR EXCLUSION FROM SELECTION

QUESTIONS

(1) Why is comparative genomics useful in the identification of potentially functional non-coding regions? What is the function of many of the conserved non-coding regions?

(2) Four founder transgenic mice (A-D) were generated by classical transgene injection into zygote and identified by tail-DNA transgene-PCR. As hoped, all of them expressed the flagged plasmamembrane protein encoded by the transgene. Upon crossing to wild type females, their F1, PCR-positive progenies showed different features: F1(A), F1(B) and F1(C) expressed the protein, F1(D) did not. F1(A-C) were further backcrossed and their transgenic descendants (again identified by PCR), displayed a further diversified expression pattern: F2(A) and F2(B) expressed the protein, F2(C) did not. We are afraid that the lines (A) and (B) might subsequently become "silent" too and, therefore, we do not feel comfortable with using them for our long term assays. Please provide possible and simple explanations for these problems and - if possible - suggest how fixing/ preventing them.

(3) The impact of genomics in the study of neurodegenerative diseases

(4) Explain the reasons why transposable elements are believed to be at the basis of the birth of many long non-coding RNAs.

(5) To a first approximation, mice generate neurons belonging to different layers of their neocortex starting from a common proliferating precursor pool, according to the sterotyped temporal sequence (from earlier to later): (1,7), 6, 5, 4, 3, 2. Gene A is suspected to be implicated in regulating the fraction of neurons adopting layer 5 identity and location. If we decrese or increase the A allele copy number, this fraction decreases. Explain molecular mechanisms by which opposite gene manipulations can give rise to the same result. Propose rationale cellular mechanisms through which an altered gene dosage may lower layer 5 neuron frequency. Suggest simple experiments to test the mechanisms proposed.

(6) Non-coding RNA in neurodegenerative diseases

(7) Analyzing sequencing data mapped on the genome from DNA of a single human postmortem brain you find that, in a unique specific genomic location, the reads mapped display the presence of 3 different nucleotides: A, T and C. The A results to be present in about 48% of the reads, the T is present in 47% of reads while the C is present only in 5% of reads mapping on that position. The coverage of the region is very high and all the reads with errors were removed; therefore these nucleotides are real and their frequencies are not an error and/or an artifact. How can you explain this result? What A, T and C represent?

(8) Viral vectors for gene therapy: key traits and choice criteria.

(9) Somatic mutations in diseases.

(10) Single cell transcriptomics approaches may be quite more expensive compared to conventional *en mass* transcriptomics. I which cases do you think the former approach provides a key advantage over the latter, so that - despite costs - it is preferable over it?

(11) Split reads and discordant reads. Explain what they are, why are they useful and which kind of variations they can identify.

(12) Pharmacogenomics

(13) Describe how transposable elements generate somatic variations in the brain.

(14) Induced pluripotent stem cells: derivation, maintenance, in vivo correlates, fields of use.

(15) The gene XYZ is transcribed from the locus xyz in a fish species living in two different places: the Indian Ocean and the Southern Ocean. The sequence of the transcript for the gene XYZ is almost identical between fishes from the two places; there is only a single nucleotide difference: in position 39 of the transcript the species from the Indian Ocean has a G while the species from the Southern Ocean has an A. However, when researchers sequence the genome for the xyz locus both the groups of fishes show to have an A in that position. How can you explain this difference at the RNA level and not in the DNA?

(16) The role of the environment in the epidemiology of neurodegenerative diseases

(17) We suspect that gene X increases the amount of protein Y available at the post-synaptic membrane, specifically by promoting translation of Y-mRNA. To assess this issue, we: (a) overexpress X, (b) measure Y-protein by WB and Y-mRNA and qRTPCR, and (c) compare the [Y-protein]/[Y-mRNA] ratio. Such ratio is upregulated in in X-overexpressing vs control samples: was our hypothesis correct? explain why. If appropriate, what would you do to definitively clarify this issue?

(18) Describe the life-cycle of autonomous and non-autonomous non-LTR retrotransposons.