MCD 2021 First Level Exam- PART ONE
Part 1: Essay Questions. ANSWER 4 QUESTIONS out of THE 6 CHOICES.

EACH ESSAY QUESTION IS WORTH 15 POINTS General Essay is opened from 9 am until 11 pm
Time yourself and make sure to submit BEFORE OR BY the deadline of 11 pm. Set a timer and submit whatever you have before or by 11 pm.

You must answer 4 of the 6 essay choices. Clearly indicate the number and the references for each of your essays. The essays must appear in numerical order in your written document. For example, if you answer choices 2, 4, 5, and 6 then they must appear in that sequential order. DO NOT COPY AND PAST ANY INFORMATION!!! YOU CAN USE THIS DOCUMENT TO ANSWER IN AND CAN DELETE THE TWO QUESTIONS THAT YOU DO NOT ANSWER.

Choice 1 Essay
References:
(Liu et al., 2020; Sato et al., 2014)

Liu, Z., Shi, H., Nzessi, A.K., Norris, A., Grant, B.D., and Liu, J. (2020). Tetraspanins TSP-12 and TSP-14 function redundantly to regulate the trafficking of the type II BMP receptor in Caenorhabditis elegans. Proc Natl Acad Sci U S A *117*, 2968-2977.

Sato, K., Norris, A., Sato, M., and Grant, B.D. (2014). C. elegans as a model for membrane traffic. WormBook, 1-47.

A (6 points). Barth Grant's lab has identified components involved in membrane trafficking and generated fusion proteins that mark different intracellular compartments in *C. elegans*. For the following markers, describe how the intracellular compartments relate to one another.

LMP-1 lysosome
RAB-5 early endosome
RAB-7 late endosome
RAB-11 recycling endosome
SYN-16 late Golgi
TRAM endoplasmic reticulum
VPS-35 retromer

B (1 point) Jun (Kelly) Liu's lab has determined that normal localization of the BMP type II receptor DAF-4 in *C. elegans* requires the tetraspanins TSP-12 and TSP-14. In the intestinal cells of normal animals, on what membrane domain(s) does DAF-4 localize?

C. (3 points) What is the evidence that TSP-12 and TSP-14 are required for normal localization of DAF-4? What membrane compartments are involved?

D. (2 points) For these experiments, what are the advantages and disadvantages of using endogenously tagged DAF-4 compared with a DAF-4::GFP multi-copy transgene?

E. (3 points) What are the consequences of trafficking the type I and type II receptors independently? What biological advantage could be conferred by this mechanism? How would this differ from the trafficking of a receptor tyrosine kinase?

Choice 2 Essay

References:
(Hjortness et al., 2018; Keedy et al., 2018)

Hjortness, M.K., Riccardi, L., Hongdusit, A., Zwart, P.H., Sankaran, B., De Vivo, M., and Fox, J.M. (2018). Evolutionarily Conserved Allosteric Communication in Protein Tyrosine Phosphatases. Biochemistry *57*, 6443-6451.

Keedy, D.A., Hill, Z.B., Biel, J.T., Kang, E., Rettenmaier, T.J., Brandao-Neto, J., Pearce, N.M., von Delft, F., Wells, J.A., and Fraser, J.S. (2018). An expanded allosteric network in PTP1B by multitemperature crystallography, fragment screening, and covalent tethering. Elife *7*.

Part A. Many small-molecule drugs work by inhibiting the function of a specific protein target in human cells, but they can do so by targeting different sites within that protein's 3D structure. What are the potential advantages of an allosteric inhibitor vs. an orthosteric (a.k.a. competitive) inhibitor?
[5 POINTS]

Part B. How were the protein samples prepared for the experiments in this study? What possible consequences might this method of preparation have for interpreting the results, relative to the wildtype protein in human cells? (Note that full-length PTP1B is 435 amino acids long.)
[5 POINTS]

Part C. What specific types of new structural biology experiments were used for this study, and what did these experiments reveal about possible allosteric mechanism in PTP1B? Be specific and detailed in your answer.
[5 POINTS]

Choice 3 Essay
References:
(Marteijn et al., 2014; Pines et al., 2018)

Marteijn, J.A., Lans, H., Vermeulen, W., and Hoeijmakers, J.H. (2014). Understanding nucleotide excision repair and its roles in cancer and ageing. Nat Rev Mol Cell Biol *15*, 465-481.

Pines, A., Dijk, M., Makowski, M., Meulenbroek, E.M., Vrouwe, M.G., van der Weegen, Y., Baltissen, M., French, P.J., van Royen, M.E., Luijsterburg, M.S.*, et al.* (2018). TRiC controls transcription resumption after UV damage by regulating Cockayne syndrome protein A. Nat Commun *9*, 1040.

Cells have developed multiple mechanism that contribute to the maintenance of genomic integrity. Chief among these processes are the mechanisms of DNA repair.

A.- Indicate what is the average number of DNA lesions that occur per cell per day, and indicate four different DNA repair mechanisms that are part of the cellular DNA damage response (**3pts**).

B.- How are DNA lesions recognized during global genome nucleotide excision repair (GG-NER) and transcription couple NER (TC-NER)? (**3pts**)

C.- In the context of TC-NER, why is the CSA-TRiC interaction functionally relevant? (**3pts**)

D.- Pines et al., suggest that DDB1 serves as an acceptor of the TRiC-bound CSA. What is the evidence? (**3pts**)

E.- What was the phenotype of CSA mutants, A160T, A205P, and D266G, identified in Cockayne Syndrome patients, and investigated by Pines, et al? (**3pts**)

Choice 4 Essay
References:
(Eraslan et al., 2019; Zou et al., 2019)

Zou, J., Huss, M., Abid, A. et al. A primer on deep learning in genomics. Nat Genet 51, 12–18 (2019).

Eraslan, G., Avsec, Ž., Gagneur, J. et al. Deep learning: new computational modelling techniques for genomics. Nat Rev Genet 20, 389–403 (2019).

As a data-driven science, genomics largely utilizes Machine Learning (ML) to capture dependencies in data and derive novel biological hypotheses. Deep Learning (DL), a subdiscipline of machine learning, uses computer implementations of artificial neural networks (ANNs) and enables researchers to model complex dependencies in the data using an artificial intelligence approach, improving predictive performance and hypothesis-driven discovery from large datasets.

Four major classes of ANNs are mainly used for machine learning applications, and each is best suitable for analysis of different types of genomic datasets:

(1). Describe in detail the differences between the classes of ANNs and the genomics applications each is best suited for;

(2). Provide an overview of how the data convolution process makes model predictions in Convolutional Neural Networks (CNNs).

Choice 5 Essay
References
(Kampinga and Craig, 2010; Rodriguez-Gonzalez et al., 2020)

Kampinga and Craig, 2010. The Hsp70 chaperone machinery: J-proteins as drivers of functional specificity. Nat Rev Mol Cell Biol, 11(8): 579–592.

Claudio Rodríguez-González et al., 2020. Co-chaperones DNAJA1 andDNAJB6 are critical for regulation of polyglutamine aggregation. Sci Rep. 2020, May 18;10(1):8130.

Chaperone partners Hsp70 and Hsp40 are found in all three domains of life and are among the most evolutionarily conserved of proteins. Humans have 11 Hsp70 genes, with a variety of overlapping subcellular and extracellular localizations, but highly conserved sequence and structure. Hsp40s are best known as regulators of Hsp70 activity; the human genome encodes >40 Hsp40 proteins.

A. Describe the molecular mechanism underlying the refolding activity of the Hsp70-Hsp40 complex. What is the function of Hsp40s that lack a peptide-binding domain? (5 pts)

B. In addition to regulating the folding activity of Hsp70s, Hsp40s have other functions. Describe the mechanisms by which Hsp40s influence Hsp70 function, as well as the proposed independent functions of Hsp40s. (5 pts)

C. Rodríguez-González and colleagues show that loss of Hsp40 proteins DNAJB6 and DNAJA1 have opposite effects on the aggregation of polyQ74htt (huntingtin exon 1 with 74 glutamines in the polyQ repeat tract). Knock-out (KO) of DNAJB6 increases aggregation, whereas KO of DNAJA1 decreases aggregation. Speculate - suggest two plausible molecular mechanisms to explain the phenotype of the DNAJB6 knock-out cells and, separately, the DNAJA1 knock-out cells. (5 pts)

Choice 6 Essay
References:
(Dufrene and Persat, 2020; Ponisch et al., 2018)

Dufrene, Y.F., and Persat, A. (2020). Mechanomicrobiology: how bacteria sense and respond to forces. Nat Rev Microbiol *18*, 227-240.

Ponisch, W., Eckenrode, K.B., Alzurqa, K., Nasrollahi, H., Weber, C., Zaburdaev, V., and Biais, N. (2018). Pili mediated intercellular forces shape heterogeneous bacterial microcolonies prior to multicellular differentiation. Sci Rep *8*, 16567.

Bacteria are the most ancient life forms to populate our planet. For the longest time, we thought that they were mostly living by free-floating in liquid environments, a planktonic lifestyle. Within the last few decades, scientists have realized that bacteria live in dense multicellular communities attached on surfaces called biofilms.

A) What type of cues can be present within biofilms that will not be present in the planktonic lifestyle? Explain how those cues could lead to the transition between the planktonic and biofilm lifestyle.

B) Present two organelles known to mediate the transition between these two lifestyles and explain potential signaling associated with them.

C) Drawing a parallel with the development of eukaryotic multicellular organisms, explain how biofilms can organize themselves.

MCD 2021 First Level Exam- PART TWO
Part 2: EACH QUESTION IS WORTH 20 POINTS

This session will be opened from 2pm until 4pm

Time yourself and make sure to submit BEFORE OR BY 4 pm. Set a timer and submit whatever you have before or by the 4 pm mark.

You must answer 1 of the 2 experimental design choices and 1 of the 2 data analysis choices. Clearly indicate the number and the references for your answers. The answers must appear in numerical order in your written document. For example, if you answer choices 7, and 10 then they must appear in that sequential order.

PART 2A: PICK ONE Experimental Design QUESTION, WORTH 20 POINTS

Choice 7. Experimental Design
References:
(Corces et al., 2016; Gasperini et al., 2020)

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., *et al.* (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat Genet *48*, 1193-1203.

Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat Rev Genet *21*, 292-310.

- Corces et al. developed a method to identify transcription factors involved in hematopoietic development and differentiation (Fig. 4 and Supplementary figure 9).

A. How do the authors narrow down the list of candidate transcription factors among those belonging to the same family? Please articulate the reason for your answer by providing specific relationships to the data presented in the paper.

B. Using the results advanced by this report, design an experimental approach that will demonstrate the involvement of an evolutionarily conserved transcription factor in mammalian hematopoiesis and identify all of the protein's direct downstream cis-regulatory regulatory elements (CREs) and target genes.

C. What are the limitations of this whole "Corces et al" approach to comprehensively identify all the transcription factors and regulators involved in hematopoiesis? Please detail some of the potential confounding possibilities.

D. The authors extend their analysis to characterize the genomic accessibility deviations during hematopoiesis associated with SNPs of disease GWAS (Fig. 4h-k).
Explain how you will use these results to facilitate your design of experimental strategies to identify and validate the CRE variants contributing to the diseases mentioned.

Choice 8. Experimental Design

References:

(Wei et al., 2016; Weissman and Gage, 2016)

Wei, P.C., Chang, A.N., Kao, J., Du, Z., Meyers, R.M., Alt, F.W., and Schwer, B. (2016). Long Neural Genes Harbor Recurrent DNA Break Clusters in Neural Stem/Progenitor Cells. Cell *164*, 644-655.

Weissman, I.L., and Gage, F.H. (2016). A Mechanism for Somatic Brain Mosaicism. Cell *164*, 593-595.

1. Wei et al published a novel hypothesis to suggest a function for recurrent DNA double-strand breaks in long neural genes, such as *Lsamp*, of neural stem/progenitor cells (NSPCs). However, the authors do not provide data to show the biological significance (e.g. neuronal diversity) of the recurring DNA breaks in these genes. Design an experiment to validate the biological significance of a gene that harbors recurring DNA double-strand breaks. Make sure to clearly articulate all the biological material and reagents you will need for this experiment. (10 points)

2. In Table S1, the authors show data to indicate that *ATM* is required for recurring low-level DNA double-strand breaks in NSPCs. The authors suggest that the loss of ATM causes a defect in oxidative stress-dependent DNA damage repair. However, the authors do not provide data to show that the ATM-dependent oxidative stress response rather than the ATM-dependent DNA damage responses is necessary for the low level DNA double-strand breaks. Design an experiment to distinguish between these two possibilities. Make sure to clearly articulate all the biological material and reagents you will need for this experiment. (10 points)

PART 2B: PICK ONE Data Interpretation QUESTION, WORTH 20 POINTS

Choice 9. Data Interpretation

References:

(Barkess and West, 2012; Nora et al., 2017; Wallace and Felsenfeld, 2007)

Barkess, G., and West, A.G. (2012). Chromatin insulator elements: establishing barriers to set heterochromatin boundaries. Epigenomics *4*, 67-80.

Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944 e922.

Wallace, J.A., and Felsenfeld, G. (2007). We gather together: insulators and genome organization. Curr Opin Genet Dev *17*, 400-407.

You are studying a region of DNA (region X) that lies in between two genes (Gene Y and Gene Z). In whole animals, Gene Y's expression pattern is very different from that displayed by Gene Z, indicating that two distinct gene expression programs are operating on either side of region X. Robust transcriptional enhancers for both Gene Y and Gene Z have been mapped and identified in their respective loci. You strongly suspect that region X is playing a key role in keeping the influences of the Gene Y and Gene Z regulatory programs separated. This would prevent the Gene Y program from interfering with the Gene Z expression program, or vice versa. In "W" cells, Gene Y is highly expressed while Gene Z is silent. Experiments are carried out to create W cells that are genetically engineered to completely delete region X DNA in the genome (XKO). Please answer the following:

A. If Gene Z becomes abnormally expressed in XKO W cells:

1) What would you hypothesize is the type of cis-acting gene regulatory activity supported by region X? Please clearly explain the rationale for your hypothesis. (5 points)

2) Briefly discuss the role of at least one key protein, and at least one key gene regulatory mechanism involved in the activity you describe in A1. (5 points)

B. If gene Y becomes abnormally silent in XKO W cells:

1) What would you hypothesize is the type of cis-acting gene regulatory activity supported by region X? Please clearly explain the rationale for your hypothesis. (5 points)

2) Briefly discuss the role of at least one key protein, and at least one key gene regulatory mechanism involved in the activity you describe in A1. (5 points)

<u>Choice 10. Data Interpretation</u>:
References
(Andersen et al., 2020; Yao et al., 2020)
Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., and Garry, R.F. (2020). The proximal origin of SARS-CoV-2. Nat Med *26*, 450-452.

Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z.*, et al.* (2020). Molecular Architecture of the SARS-CoV-2 Virus. Cell *183*, 730-738 e713.

20 points total

Background: The year 2020 saw the emergence of the SARS-CoV-2 coronavirus responsible for the COVID-19 pandemic. This is the seventh coronavirus known to infect humans and pathogens of this group are believed to originate from other mammals through zoonotic transfer. Assume that a severe respiratory infection of domestic cattle has been identified on a large industrial ranch in the western United States. Health authorities need to rapidly identify the virus to determine if it is a newly emerged coronavirus that may have transferred to humans or perhaps a transfer of the human virus to bovine hosts.

Materials: The major reagents available to work with are (1) nasal washes from several infected cattle (2) African green monkey kidney (VERO) cells known to be permissive for coronavirus replication with mammalian culture medium including fetal bovine serum (3) Phenol:Chloroform:Isoamyl Alcohol and nonionic detergent (4) commercial cDNA synthesis kit with reverse transcriptase for first strand DNA polymerization with random primers (5) access to a commercial nucleic acid sequencing facility and nucleic acid genome viral database (6) biosafety level laboratory (BSL-3).

Questions:

a. (5 pts) Describe the method the researchers would have used to isolate and purify viral particles suitable for genome isolation.

b. (5 pts) How did they demonstrate that the pathogen contains an RNA genome?

c. (10 pts) What relevant information can be deduced from the comparison and alignment with the spike protein amino acid sequence as compared to known coronaviruses in regard to a closest relation and functional regions? Refer to the annotated alignment image below: I.e. new isolate “a” “b” (Andersen KG. *et.al.*, *Nature Medicine, March, 2020)*.