MCB C243
Proposal for a new graduate course
Genetics, Genomics, & Development Division,
Department Molecular & Cell Biology
(planned crosslisting as Math C243)

submitted by Lior S. Pachter,
Professor, Math & MCB.

Metadata for the course:

MCB C243/Math C243
Course title: *Seq: Methods and Applications
(*Note: "*Seq" is pronounced "Star Seq".*)
Course description: A graduate seminar class in which a group of students will closely examine recent computational methods in high-throughput sequencing followed by directly examining interesting biological applications thereof.
Units: 3
Offered: Spring
Format: Lecture only
TuTh Lecture
Time: 8:00am-9:30am
Grading: Letter
Final Exam: No scheduled final exam-Final Project
Cross-listed: Math Department-Yes.
Courses that will restrict credit: None.
Prerequisites: graduate standing in Math, MCB, and Computational Biology; or consent of the instructor.
Repeatable for credit: No
Number of students expected to enroll: 8 - 24 students.

Detailed Proposal

The proposal is to create MCB C243 "*Seq: Methods and Applications" as a graduate seminar class in which a group of students will closely examine recent computational methods in high-throughput sequencing followed by directly examining interesting biological applications thereof. The class will meet twice a week for sessions of 1.5 hours. For the first two weeks, the professor will review the basic biology and computational/mathematical techniques underlying high throughput sequencing as a biological tool. Students will also get a chance to choose an application paper to present for weeks 2- 8. The next six weeks will start by focusing on a seminal method paper or an exceptionally clear review article, for which the professor will lead students in an exposition of the methods, provide a clarifying summary and answer questions. Each Thursday a student or students will present an application of the week's method and

lead a group discussion about the application for the second session. (The selection of papers presented below for the "Thursday menus" is merely representative, not exhaustive.) The 9th week will be devoted to an extended discussion of what methods and applications the students found particularly compelling or insightful, their proposed final project, and choosing a paper relating to their final project they would like to present. Students will submit a final project in consultation with the professor, choosing from several options a) writing a review article of a method not reviewed in week 1-8 b) writing a review article of several applications of a method to a specific disease or problem in biology c) their own a their own attempt at running a computational analysis based on any *Seq method paper or c) a paper motivating and sketching out possible strategies and pitfalls of a new method. 8 - 24 students. (8-16 presentations first half of class, 12 - 24 presentations second half of class.)

30% Productive class participation
20% first presentation of a paper
20% second presentation of a paper
10% Insight/Planning Presentations & Discussion (Week 9)
20% Final project.

Sample syllabus (exact methods may differ from year-to-year):

Week 1: A review of high-through put sequencing technology. Students sign up for the Week 2-8 presentations and choose the application paper they wish to present.

Week 2: A review of basic mathematical and computational ideas in analyzing sequencing data. Pass/Fail problem set.

Week 3 - 8 Students will be expected to read the method paper before lecture on Tuesday so they can ask questions about it. They will be expected to read the application paper their colleague has chosen for the week before Thursday.

Week 3: DNA-ase seq:

Tuesday: Gregory E. Crawford et al., "Genome-wide Mapping of DNase Hypersensitive Sites Using Massively Parallel Signature Sequencing (MPSS)," *Genome Research* 16, no. 1 (January 1, 2006): 123–131, doi:10.1101/gr.4074106.

Thursday: Students' choice:a) G-quadruplexes in promoters throughout the human genome, Huppert & Balakrishnan. Nucl. Acids Res. (2007) 35 (2): 406-413. doi: 10.1093/nar/gkl1057; b) Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, et al. (2006) Retroviral DNA Integration: Viral and Cellular Determinants of Target-Site Selection. PLoS Pathog 2(6): e60. doi:10.1371/journal.ppat.0020060 c)

Week 4: CHIP-seq

Tuesday: David S. Johnson et al., "Genome-Wide Mapping of in Vivo Protein-DNA Interactions," *Science* 316, no. 5830 (June 8, 2007): 1497–1502, doi:10.1126/science.1141319.

Thursday: Students' choice: a) Science 21 May 2010: Vol. 328 no. 5981 pp. 1036-1040 DOI: 10.1126/science.1186176 b) A ChIP-seq defined genome-wide map of vitamin D receptor binding: Associations with disease and evolution Ramagopalan et al, August 24, 2010, doi: 10.1101/gr.107920.110 Genome Res. 2010. 20: 1352-1360

Week 5: RAD-seq

Tuesday: Nathan A. Baird et al., "Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers," *PLoS ONE* 3, no. 10 (October 13, 2008): e3376, doi:10.1371/journal.pone.0003376

Thursday: Students' Choice a) Resolving postglacial phylogeography using high-throughput sequencing, Emerson et al doi: 10.1073/pnas.1006538107 PNAS September 14, 2010 vol. 107 no. 37 16196-16200 b) Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, et al. (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. PLoS Genet 6(2): e1000862. doi:10.1371/journal.pgen.1000862

Week 6: RNA-SEQ I:

Tuesday: Ali Mortazavi et al., "Mapping and Quantifying Mammalian Transcriptomes by RNA-Seq," *Nature Methods* 5, no. 7 (July 2008): 621–628, doi:10.1038/nmeth.1226.

Thursday: Students' Choice a) New insights into the blood-stage transcriptome of Plasmodium falciparum using RNA-Seq. Otto TD, et al. Mol Microbiol. 2010 Apr;76(1):12-24. DOI: 10.1111/j.1365-2958.2009.07026.x b) Mammalian Genome Volume 21, Issue 11-12 , pp 592-598 DOI 10.1007/s00335-010-9297-z c) Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq, Bruno et al. doi: 10.1101/gr.109553.110 Genome Res. 2010. 20: 1451-1458

Week 7: RNA-SEQ II:

Tuesday: **Cole Trapnell**, Lior Pachter, and Steven L. Salzberg, "TopHat: discovering splice junctions with RNA-Seq." *Bioinformatics* (2009), doi:10.1093/bioinformatics/btp120

Thursday: Student's Choice: a) Twine NA, Janitz K, Wilkins MR, Janitz M (2011) Whole Transcriptome Sequencing Reveals Gene Expression and Splicing Differences in Brain Regions Affected by Alzheimer's Disease. PLoS ONE 6(1): e16266. doi:10.1371/journal.pone.0016266 b) Directional gene expression and antisense transcripts in sexual and asexual stages of Plasmodium falciparum, María J López-Barragán et al, *BMC Genomics* 2011, **12**:587 doi:10.1186/1471-2164-12-587

Week 8: Methyl-Seq

Tuesday: Alayne L. Brunner et al., "Distinct DNA Methylation Patterns Characterize Differentiated Human Embryonic Stem Cells and Developing Human Fetal Liver," *Genome Research* 19, no. 6 (June 1, 2009): 1044–1056, doi:10.1101/gr.088773.108.

Thursday: Students' choice.

Week 9: Insight and Planning presentations (5-15 minute presentations on what they found most insightful week 1-8, what they would like to do for their final project, and therefore what paper they would like present week 10-14).

Week 10 - 14: Students present on a paper relating to their final project.

Week 15: Students present their final project.