

## Bio 4342, Research Explorations in Genomics

Draft, Spring 2016

M, W 1:30-5:00; F 1:30-2:30 (occasionally 3:30), Writing Intensive group F 2:30-3:30 when scheduled. Meet in LS 311, Danforth Campus; on Friday 1/22 there will be a visit to the Washington University McDonnell Genome Institute. Please review our research problem (on the course website at <http://www.nslc.wustl.edu/courses/Bio4342/bio4342.html>) and read "A Guide to Consed" on the GEP website at [http://gеп.wustl.edu/curriculum/course\\_materials\\_WU/sequence\\_improvement/tutorials\\_and\\_walkthroughs](http://gеп.wustl.edu/curriculum/course_materials_WU/sequence_improvement/tutorials_and_walkthroughs)) prior to the first class.

- 1/20 Wed **Course structure**; research problem overview; assembly challenge (Elgin, 45 min).  
**Lab**: Intro to computers; UNIX commands to get you started; begin work on *Using Consed Graphically* (navigation; adding reads) (Shaffer, Trani).  
**Lab**: Pre-course assessment, GEP survey and quiz (45 min) (<http://gеп.wustl.edu>).  
**Reading**: Please watch Genome Center Virtual Tour and Next Generation Sequencing Video Tour on line prior to class on Friday.  
[http://gеп.wustl.edu/curriculum/course\\_materials\\_WU/introduction\\_to\\_genomics/](http://gеп.wustl.edu/curriculum/course_materials_WU/introduction_to_genomics/)  
The **associated worksheet** (DiAngelo+SCRE, in folder) will be collected at the end of the tour on Friday.
- 1/22 Fri **Tour** of the Genome Institute (be sure to wear long pants, closed shoes) (Cherilynn Shadding, TGI staff)  
NOTE: meet in the elevator lobby on the Fourth Floor, 4444 Forest Park Parkway (1:14 pm train from Skinker to Central West End Metro stop).
- 1/25 Mon **Lecture**: Overview of DNA sequencing (goals); the pipeline; different sequencing strategies (**E Mardis**, 45 min plus discussion)  
**Lecture**: The basics on finishing *D. ficusphila* hybrid assemblies (30 min) (Shaffer)  
**Lab**: Complete *Using Consed Graphically*; start *GEP Hybrid Assembly Walkthrough* (using navigators, assessing quality, making corrections) (Shaffer, Trani). Assign HW#1 (Consed)  
**Reading**: "Sleeping dogs of the genome," Gorbunova et al; start Ellison & Bachtrog + commentary (RR due 1/29).
- 1/27 Wed **Lecture**: The challenges in generating finished sequence; questions on HW #1 (Shaffer).  
**Lab**: Continue *Finishing a Drosophila Hybrid Assembly* (corrections, resolving gaps, PCR primer design) (Trani, Shaffer)  
Introduce reference materials and finishing report requirements (including *Finishing Checklist*), obtain sequence file for own *D. ficusphila* project, begin analysis (Leung; Trani, Shaffer).

- 1/29 Fri **Discussion:** Ellison & Bachtrog; **RR#1 due** (Elgin, ~60 min.)  
**WI subgroup: 1000 word paper due** (extended RR#1); discussion on scientific writing, peer review (~30 min) (Bednarski, SCRE).  
**Optional Lab** (~1 hr): Work with Consed, HW1
- 2/1 Mon **Lab: HW#1 due.** Finishing own project with help from TGI finishers available. Discussion of Miniassembly, Consed, and assessment of sequence quality as needed. Review of when to request additional sequencing, designing primers, as needed (Shaffer).  
**Suggested reading:** Figures in Treangen & Salzberg, 2012.  
**Presentation:** Bio 4342 alumni present prototypes of oral finishing report.
- 2/3 Wed **Lab meeting: 9 min presentation** + 3 discussion each (individual reports in groups of five, one group starts at 1 pm) "Finishing my project; problems identified and solved; remaining issues." Continue to finish own project based on feedback.
- 2/5 Fri **Lecture:** Eukaryotic genomes/chromatin structure (Elgin, 45' + discussion)  
Optional help session (~1 hr).  
**WI subgroup: Critique** of 1000 word paper due; discussion.  
**Reading:** Eddy (2012) The C-value paradox....; and the modENCODE page on chromatin, including the vignette on fly chromatin  
(<http://modencode.sciencemag.org/chromatin/introduction> )  
**Optional homework:** Cot curve analysis.
- 2/8 Mon **Lecture:** Heterochromatin/euchromatin (Elgin, 45' + discussion).  
**Lab:** Finishing own project (TGI staff); checklist for project submission; final day for consultation  
**Reading:** Start Haynes et al. (RR#2 due 2/13).
- 2/10 Wed **Lecture:** Chromatin states (modENCODE); dot chromosome, GEP findings (Elgin)  
**Lecture/Demo:** Introducing BLAST (Yu He).  
**Lab:** Simple Introduction to BLAST (scripted walk-through).  
**Finishing papers due.** Submit data files. (WI subgroup: exchange finishing papers)
- 2/12 Fri **Lecture:** Introduction to web databases and FlyBase (Leung)  
**WI subgroup: Rewrite** of 1000 word paper due.  
**Reading:** Webber & Ponting, 2004
- 2/15 Mon **Lecture:** Gene finding: detecting and interpreting genetic homology (Buhler)  
**Lab:** Begin CS HW1.  
**Optional homework** on Cot curves due.
- 2/17 Wed **Lecture:** Browser-Based Annotation and RNA-Seq Data (Buhler)  
**Lab:** Begin CS HW2. **CS HW1 due** by end of class.  
**Guest lecture:** Ting Wang: Insights from the human epigenome browser

- 2/19 Fri **Discussion:** Haynes et al, Yandim et al (Elgin) **RR#2 due.**  
**WI subgroup: peer review** of finishing papers due.
- 2/22 Mon **Lecture:** Introduction to gene predictors (Shaffer).  
**Lab:** Gene prediction tutorial (train with chimp Contig95, Genscan); organize chimp project teams. **CS HW2 due.**
- 2/24 Wed **Guest lecture: John Edwards: The role of DNA methylation**  
**Lab:** continue with chimp gene finding lab with partners.  
**Presentation:** Bio 4342 alumni present chimp genome oral reports
- 2/26 Fri **Presentation:** Bio 4342 alumni present chimp genome oral reports  
**Lab:** Prepare lab meeting presentation with partners  
**WI subgroup: Rewrite** of finishing papers due.
- 2/29 Mon **Lab:** Annotation of fragments of the chimp genome.  
**Lab meeting: oral presentations** (12 min each group, starting ~3 pm) “Genes and pseudogenes in chimp.”
- 3/2 Wed **Lecture:** Introduction to *ab initio* and evidence-based gene finding (Leung)  
**Lab:** Complete work on chimp annotation.
- 3/4 Fri **Lecture:** Hidden Markov Models (Yu He)  
**Chimp analysis paper due** (WI exchange papers)  
**Reading:** Eddy, S (2004a) What is a Hidden Markov Model?
- 3/7 Mon **Lecture:** Cancer genetics (E Mardis)  
**Reading:** Wartman (2015) A case of me.  
**Lab:** Hidden Markov Models exercise (CS HW3)  
**Discussion:** start Leung et al (RR #3 due 3/25); overview discussion (Elgin, 20’)
- 3/9 Wed **Lecture:** Dynamic programming (similarity searches) (Yu He)  
**Lab:** Begin Dynamic programming exercise (CS HW4)  
**Reading:** Eddy, S (2004b) What is dynamic programming?  
**WI subgroup: peer review** of Chimp paper due, discussion.
- 3/11 Fri Optional help session  
**CS HW3 due.**
- 3/14 - 3/18 **Washington University Spring Break**

- 3/21 Mon **Lecture:** An introduction to the annotation projects (Elgin, 20')  
**Lecture:** Efficient gene finding in Drosophila (Shaffer)  
**Lab:** Annotation of a Drosophila gene (scripted walk-through).  
**CS HW4 due.**
- 3/23 Wed **Lab:** claim projects (Leung); begin work with own Drosophila project to identify genes (Leung, Shaffer, TAs).
- 3/25 Fri **Discussion:** Leung et al 2015 (Elgin, Leung); **RR #3 due**
- 3/28 Mon **Lecture:** Review of gene model checker (Leung)  
**Lab:** Gene annotation; check first gene model by end of class.  
**WI subgroup: Rewrite** of chimp analysis due
- 3/30 Wed **Lecture:** Searching for transcription start sites (Leung)  
**Lab:** Continuing annotation project; check for regulatory elements of first gene
- 4/1 Fri **Lecture:** Motif finding (**J Buhler**); motif hunting exercise (begin CS HW5)
- 4/4 Mon **Lab Meeting:** **oral report** on first gene (9' + 3' discussion; individual reports in groups of five, one group starts at 1 pm) "Annotation of the first gene in my project"
- 4/6 Wed **Lecture:** RepeatMasker; other ways to find repeats (**Buhler**).  
**CS HW5 due.**
- 4/8 Fri **Lecture:** Finding repeats in Drosophila (Leung)  
**Draft report on one gene due** (WI exchange papers)
- 4/11 Mon **Lecture:** Targeting heterochromatin formation in Drosophila (Elgin)  
**Lab:** Continue analysis and annotation of own project.  
**Reading:** start #4 (motif hunting, regulation); RR #4 due 4/22
- 4/13 Wed **Lecture:** Further characterization of genes/gene products: use of Clustal, use of FlyBase resources (**Yu He**)  
**Lab:** Preparation of annotation reports; gene intron/exon structures  
**Presentation:** Bio 4342 alumni present annotation project oral reports
- 4/15 Fri **Guest lecture:** Rick Wilson, Human Genetics  
**WI subgroup: critiques** of annotation drafts due, discussion.
- 4/18 Mon **Guest lecture:** The RNAi system  
**Lab:** Preparation of annotation reports; TSS determinations; functions of genes
- 4/20 Wed **Lab:** Preparation of annotation reports; TSS determinations; functions of genes  
**Lab Meeting:** **oral report** on finding TSSs (9' + 3' discussion; individual reports in

groups of five, first group at 1 pm)

- 4/22 Fri **Discussion:** Paper #4 (RR #4 due).
- 4/25 Mon **Lecture:** Silencing repeats in *Drosophila* (Elgin)  
**Lab:** Preparation of annotation reports; start PPT preparation.
- 4/27 Wed **Lab:** Preparation of annotation reports
- 4/29 Fri **Lab:** Final consultations on annotation report
- 5/2 Mon **Final written and oral annotation reports:** Submit final paper on your project, with a map of genes (including estimates of transcription start sites), repetitious elements, and alignment to *D. melanogaster*, including a discussion of synteny. Complete annotation of all exons, all isoforms. Include results of searches for TSS candidate sites and regulatory elements. As time permits, exploration of one gene on FlyBase, expanding on gene features, regulation, and function. Use Clustal at least once. 10' presentations (1 pm 3 pm OR 3 pm – 5 pm in LS 311).
- 5/3 Tue **Course Assessment:** Follow-up session on course evaluation, submission of final files, return of computers, etc. (12 noon lunch – 2 pm, LS 311)  
(<http://evals.wustl.edu>; GEP web site; and Bio 4342 surveys/suggestions).

## Reading

We will read and discuss four papers over the course of the semester, centered on the theme of genome organization and evolution in *Drosophila*, with an emphasis on the role of repetitious elements. [If you have not read scientific papers before, look at “How to Read a Scientific Paper” by Mary Williams, pp 1-5 (on the Bio 4342 website) before starting.] These papers are listed below; for each paper you will turn in a “reading reflection” (~2 pages, double-spaced, typed) that summarizes the big idea and proposes the next experiment. In addition, we have assembled a list of papers that are pertinent to the material we will be discussing, including papers recommended by our guest lecturers. Among these, papers marked “R” are highly recommended background reading. Background material on BLAST and other computer programs can be found in the recommended texts, and on-line through our subscription to “Current Protocols in Bioinformatics,” available at <http://onlinelibrary.wiley.com/book/10.1002/0471250953>. Background information on many scientific terms is available through the Genomics Education Partnership Glossary (<http://gep.wustl.edu>) and information on terms and techniques is available through Wikipedia (generally a good source, but be cautious!).

### Required Reading (paper copies provided; reading reflections due):

1. Ellison CE, Bachtrog D. (2013) Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* 342: 846-50. (See also Chuong EB & Feschotte C (2013) Evolution: Transposons up the dosage. *Science* 342: 812-13.)
2. Haynes KA, Caudy AA, Collins L, Elgin SCR (2007) Element 1360 and RNAi components

contribute to HP1-dependent silencing of a pericentric reporter. *Curr Biol* 16: 2222-7. (See also Grewal & Elgin (2002) for background on concepts tested in this paper.)

3. Leung W et al. (2015) Muller F elements maintain a distinct set of genomic properties over 40 million years of evolution. *G3 Genes|Genomes|Genetics* 5: 719-40. Focus your review and experiment either on genome organization (figures 1-4, 9) OR on properties of genes (figures 5-8).

4. Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ, Kuroda MI (2008) A sequence motif within chromatin entry sites directs MSL establishment on the Drosophila X chromosome. *Cell* 134: 599-609.

### **Additional References:**

### **Sequencing Technology**

Before class starts:

“Sequencing a Genome” (view on line, including review of chemistry if needed):

[http://gep.wustl.edu/curriculum/course\\_materials\\_WU/introduction\\_to\\_genomics/tour/html/gsc.htm](http://gep.wustl.edu/curriculum/course_materials_WU/introduction_to_genomics/tour/html/gsc.htm) **R**

“Next generation Sequencing: Genome Center Video Tour” (view all four segments online):

[http://gep.wustl.edu/curriculum/course\\_materials\\_WU/introduction\\_to\\_genomics/nextgen\\_video\\_tour](http://gep.wustl.edu/curriculum/course_materials_WU/introduction_to_genomics/nextgen_video_tour) **R**

Weeks 1-3: Heather JM, Chain B. (2015) The sequence of sequencers: The history of sequencing DNA. *Genomics*. 2015 Nov 10. pii: S0888-7543(15)30041-0. doi: 10.1016/j.ygeno.2015.11.003.

Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. (2013) The next-generation sequencing revolution and its impact on genomics. *Cell* 155: 27-38. doi: 10.1016/j.cell.2013.09.006.

Miyamoto M, Motooka D, Gotoh K, Imai T, Yoshitake K, Goto N, Iida T, Yasunaga T, Horii T, Arakawa K, Kasahara M, Nakamura S. (2014) Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes. *BMC Genomics* 15:699. doi: 10.1186/1471-2164-15-699.

Christensen KD, Dukhovny D, Siebert U, Green RC. (2015) Assessing the Costs and Cost-Effectiveness of Genomic Sequencing. *J Pers Med*. 5:470-86. doi: 10.3390/jpm5040470.

Gordon D, Green P. (2013) Consed: a graphical editor for next-generation sequencing. *Bioinformatics*. 29: 2936-7. doi: 10.1093/bioinformatics/btt515

Nielsen, CB, Cantor, C, Dubchak, I, Gordon, D & Wang, T. (2010) Visualizing genomes: techniques and challenges. *Nature Methods* 7, S5 - S15. [Covers Consed, UCSC Genome Browser and Vist.]

Pavlopoulos GA, Malliarakis D, Papanikolaou N, Thiodosiou T, Enright AJ, Iliopoulos I. (2015) Visualizing genome and systems biology: technologies, tools, implementation techniques and trends, past, present and future. *GigaScience* 4: 38. doi: 10.1186/s13742-015-0077-2 [Recent comprehensive list.]

Treangen, TJ, and Salzberg, SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Rev. Genet.* 13: 36-46. [Nice visualization of assembly problems associated with repeats.]

## Chromatin Structure / Epigenetics

Weeks 3-5 modENCODE Consortium (2014) Chromatin, plus Vignette: Fly Chromatin. **R** <http://modencode.sciencemag.org/chromatin/introduction>

C<sub>0</sub>t curve packet with HW2. From “Biochemistry: A Problems Approach,” 2<sup>nd</sup> ed., by WB Wood, JH Wilson, RM Benbow, LE Hood; Benjamin/Cummings, CA. 1981.

Eddy, S R (2012) The C-value paradox, junk DNA and ENCODE. *Curr Biol* 22: R898-9. **R**

Palazzo AF, Gregory TR. (2014) The case for junk DNA. *PLoS Genet.* 10: e1004351. doi: 10.1371/journal.pgen.1004351.

Felsenfeld G, Groudine M (2003) Controlling the double helix, *Nature* 421: 448-453.

Li G, Reinberg D. (2011) Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev.* 21: 175-86. doi: 10.1016/j.gde.2011.01.022.

Grewal SIS, Elgin, SCR. (2002) Heterochromatin: new possibilities for the inheritance of structure, *Curr Opin Genetics & Develop.* 12: 178-187. (**R** if needed; overlaps with reading #2.)

Elgin SCR, Reuter G (2013) Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb Perspect Biol* 5: a017780. doi: 10.1101/cshperspect.a017780.

Yandim C, Natisvili T, Festenstein R. (2013) Gene regulation and epigenetics in Friedreich’s ataxia. *J Neurochem.* 126 Suppl 1: 21-42. (This paper includes a review of background information as well as recent results in a mammalian system.)

Kharchenko, PV,.... M Kellis, SCR Elgin, MI Kuroda, V Pirrotta, G Karpen, PJ Park. (2011) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 471: 480-5.

Riddle NC.....Karpen GH, Park PJ, Elgin, SCR (2012) Enrichment of HP1a on *Drosophila* chromosome 4 genes creates an alternate chromatin structure critical for regulation in this heterochromatic domain. *PLoS Genet.* 8:e1002954.

Sentmanat MF, Elgin SCR (2012) Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. Proc Natl Acad Sci USA 109: 14104-9.

Dumesic PA, Madhani HD (2014) Recognizing the enemy within: licensing RNA-guided genome defense. Trds Biochem Sci 39: 25-34.

## Human Genomics

Weeks 5+ Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, Burgess SM, Brachmann RK, Haussler D (2007) Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proc Natl Acad Sci U S A 104: 18613-18618.

Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, Zhou X, Lee HJ, Maire CL, Ligon KL, Gascard P, Sigaroudinia M, Tlsty TD, Kadlecsek T, Weiss A, O'Geen H, Farnham PJ, Madden PA, Mungall AJ, Tam A, Kamoh B, Cho S, Moore R, Hirst M, Marra MA, Costello JF, Wang T. (2013) DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nat Genet. 45: 836-41. doi: 10.1038/ng.2649.

Maunakea AK, Nagarajan RP, Bilenkyh M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielson C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SSJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466: 253-257.

Variation 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. (2015) A global reference for human genetic variation. Nature 526: 68-74. doi: 10.1038/nature15393.

Sudmant PH, et al. (2015) An integrated map of structural variation in 2,504 human genomes. Nature 526: 75-81. doi: 10.1038/nature15394.

General correlations Chen, R. et al. (2012) Personal omics profiling reveals dynamic molecular and medical phenotypes. Cell 148: 1293 – 1307.

Cancer Wartman, LD (2015) A case of me: clinical cancer sequencing and the future of precision medicine. Cold Spring Harb Mol Case Stud 1: a000349. Doi: 10.1101/mcs.a000349.

Kico, JM. et al. (2015). Association between mutation clearance after induction therapy and outcomes in acute myeloid leukemia. J Amer Med Asso 314: 811-22.

White BS, DiPersio JF. (2014) Genomic tools in acute myeloid leukemia: From the bench to the bedside. Cancer 120: 1134-44. doi: 10.1002/cncr.28552



The Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumors. *Nature* 490: 61 – 70.

Mardis ER (2014) Sequencing the AML genome, transcriptome, and epigenome. *Sem Hematology* 51: 250-58.

Carreno, BM et al. (2015) Cancer immunotherapy: A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* 348: 803-8.

Griffith M. et al. (2015) Genome Modeling System: A Knowledge Management Platform for Genomics. *PLoS Comput Biol.* 11: e1004274.  
doi: 10.1371/journal.pcbi.1004274.

DNA mC Edwards JR, O'Donnell AH, Rollins RA, Peckham HE, Lee C, Milekic MH, Chanrion B, Fu Y, Su T, Hibshoosh H, Gingrich JA, Haghighi F, Nutter R, Bestor TH. (2010) Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Res* 20: 972-80.

### Finding Genes in *Drosophila*

Weeks 8+ Hardison RC. (2003) Primer: Comparative Genomics. *PloS Biology* 1: 156-160.

Webber C, Ponting CP. (2004) Genes and homology. *Curr Biol* 14: R332-R333. **R**

Peter McQuilton, Susan E. St. Pierre, Jim Thurmond, and the FlyBase Consortium. (2011) FlyBase 101 – the basics of navigating FlyBase. *Nuc Acids Res* 39: 21.

St Pierre SE, Ponting L, Stefancsik R, McQuilton P; FlyBase Consortium (2014) FlyBase 102--advanced approaches to interrogating FlyBase. *Nucleic Acids Res.* 42: D780-8.

dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM; the FlyBase Consortium. (2015). FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.* 43(Database issue):D690-7.  
Doi: 10.1093/nar/gku1099.

Kondrashov AS. (2005) Evolutionary biology: fruit fly genome is not junk. *Nature* 437:1106. **R**

Birney E. (2007) Come fly with us. *Nature* 450: 5-6. **(Synopsis of 12 genomes paper; R)**

Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450: 203-218.

Eddy, S. (2004a) What is a hidden Markov model? *Nature Biotech.* 22: 1315-16. **R**

- Eddy, S. (2004b) What is dynamic programming? *Nature Biotech.* 22: 909-910. **R**
- Brent MR (2008). Steady progress and recent breakthroughs in the accuracy of automated genome annotation. *Nat Rev Genet.* 9: 62-73.
- Shiryev SA, Papadopoulos JS, Schäffer AA, Agarwala R. (2007). Improved BLAST searches using longer words for protein seeding. *Bioinformatics.* 23(21):2949-51.
- W. James Kent, Charles W. Sugnet, Terrence S. Furey, Krishna M. Roskin, Tom H. Pringle, Alan M. Zahler, and David Haussler. (2002) The Human Genome Browser at UCSC. *Genome Res.* 12: 996-1006.
- Rosenbloom KR, *et al.* (2015) The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* 43 (Database issue):D670-81. doi: 10.1093/nar/gku1177.
- Chen ZX, ..., Celniker SE, Oliver B, Richards S. (2014). Comparative validation of the *D. melanogaster* modENCODE transcriptome annotation. *Genome Res.* 24(7):1209-23.
- Hoskins, R. A., Landolin, J. M., Brown, J. B., Sandler, J. E., Takahashi, H., Lassmann, T., ... Celniker, S. E. (2011). Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Research*, 21: 182–192. doi:10.1101/gr.112466.110
- Brown JB, ..., Kaufman TC, Lai EC, Oliver B, Perrimon N, Graveley BR, Celniker SE. (2014). Diversity and dynamics of the *Drosophila* transcriptome. *Nature.* 512(7515):393-399.
- Lenhard B, Sandelin A, Carninci P. (2012) Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet.* Mar 6;13(4):233-45. doi: 10.1038/nrg3163.
- Palmieri N, Nolte V, Suvorov A, Kosiol C, Schlötterer C. (2012). Evaluation of different reference based annotation strategies using RNA-Seq - A case study in *Drosophila pseudoobscura*. *PLoS One.* 7(10): e46415
- Irwin Jungreis, Michael F. Lin, Rebecca Spokony, Clara S. Chan, Nicolas Negre, Alec Victorsen, Kevin P. White, and Manolis Kellis. (2011) Evidence of abundant stop codon read-through in *Drosophila* and other metazoa. *Genome Res.* 21: 2096-2113.
- Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. (2005) Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput Biol.* 1:166-75.
- Bergman CM, Quesneville H. (2007). Discovering and detecting transposable elements in genome sequences. *Brief Bioinform.* 8:382-92
- Leung, W, CD Shaffer, T Cordonnier, J Wong, MS Itano, EE Slawson-Tempel, E Kellmann, DM Desruisseau, C Cain, R Carrasquillo, TM Chusak, K Falkowska,

KD Grim, R Guan, J Honeybourne, S Khan, L Lo, R McGaha, J Plunkett, JM Richner, R Richt, L Sabin, A Shah, A Sharma, S Singhal, F Song, C Swope, CB Wilen, J Buhler, ER Mardis, SCR Elgin (2010) "Evolution of a distinct genomic domain in *Drosophila*: Comparative analysis of the dot chromosome in *Drosophila melanogaster* and *Drosophila virilis*." *Genetics* 185: 1519-1534.

Kadonaga JT. (2012). Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip Rev Dev Biol.* 1(1):40-51.

Gallo SM, Gerrard DT, Miner D, Simich M, Des Soye B, Bergman CM, Halfon MS (2011) REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in *Drosophila*. *Nucleic Acids Res.* 39: D118-23.

Zhu LJ, Christensen RG, Kazemian M, Hull CJ, Enuameh MS, Basciotta MD, Brasefield JA, Zhu C, Asriyan Y, Lapointe DS, Sinha S, Wolfe SA, Brodsky MH. (2011) FlyFactorSurvey: a database of *Drosophila* transcription factor binding specificities determined using the bacterial one-hybrid system. *Nucleic Acids Res.* 39: D111-7.

Bailey TL, Johnson J, Grant CE, Noble WS. (2015) The MEME Suite. *Nucleic Acids Res.* 2015 Jul 1;43(W1):W39-49. doi: 10.1093/nar/gkv416.

Ihuegbu NE, Stormo GD, Buhler J (2012) Fast, sensitive discovery of conserved genome-wide motifs. *J Comput Biol* 19: 139 – 47.

The following papers may be helpful during the second half of the course:

Celniker SE, Rubin GM. (2003) The *Drosophila melanogaster* genome. *Annu Rev Genomics Hum Genet.* 4: 89-117.

Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, Ashburner M, Celniker SE. (2002) The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol.* 3:RESEARCH0084. PMID: 1253757

Bartolome C, Maside X, Charlesworth B. (2002) On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol Biol Evol.* 19: 926-37.

Yang H-P, Hung T-L, You T-L, Yang T-H. (2006) Genome-wide comparative analysis of the highly abundant transposable element *DINE-1* suggests a recent transpositional burst in *Drosophila yakuba*. *Genetics*, 173: 189-96.

Thomas J, Vadnagara K, Pritham EJ. (2014) DINE-1, the highest copy number repeats in *Drosophila melanogaster* are non-autonomous endonuclease-encoding rolling-circle transposable elements (Helentrons). *Mob DNA* 5:18. doi: 10.1186/1759-8753-5-18.

Smith CD, Shu S, Mungall CJ, Karpen GH. (2007) The Release 5.1 annotation of *Drosophila melanogaster* heterochromatin. *Science* 316: 1586-91.

Hoskins RA, Carlson JW, Kennedy C, Acevedo D, Evans-Holm M, Frise E, Wan KH, Park S, Mendez-Lago M, Rossi F, Villasante A, Dimitri P, Karpen GH, Celniker SE. (2007) Sequence finishing and mapping of *Drosophila melanogaster* heterochromatin. Science 316: 1625-8.

Revised 01/12/2016