Michael S. Lawrence, PhD

Cancer results from alterations to DNA that lead to the activation of oncogenes or the inactivation of tumor suppressors. **The Lawrence laboratory** focuses on understanding the many ways this can happen, using computation as a powerful microscope to study the processes of DNA damage and repair, gene expression and genome replication, and cancer driver genes. Over our lifetimes, DNA slowly accumulates mutations due to environmental toxins and radiation, as well as from naturally occurring copying errors. The vast majority of mutations have little or no effect on a cell, but out of all possible mutations, a few may hit exactly the right place in the genome, where they can act as a "driver mutation," pushing the cell toward aggressive growth and tumor formation. Sequencing the DNA in a tumor reveals not only its driver mutations, but also all the other "passenger mutations" that were present in the tumor-initiating cell. We seek insights about cancer from both driver and passenger mutations.

Lawrence Laboratory

Andrew Dunford, BSc*
Nicholas Haradhvala, BA*
Julian Hess, BS*
Adam Langenbucher, BS
Michael S. Lawrence, PhD
Rob Morris, PhD
Ben Wittner, PhD

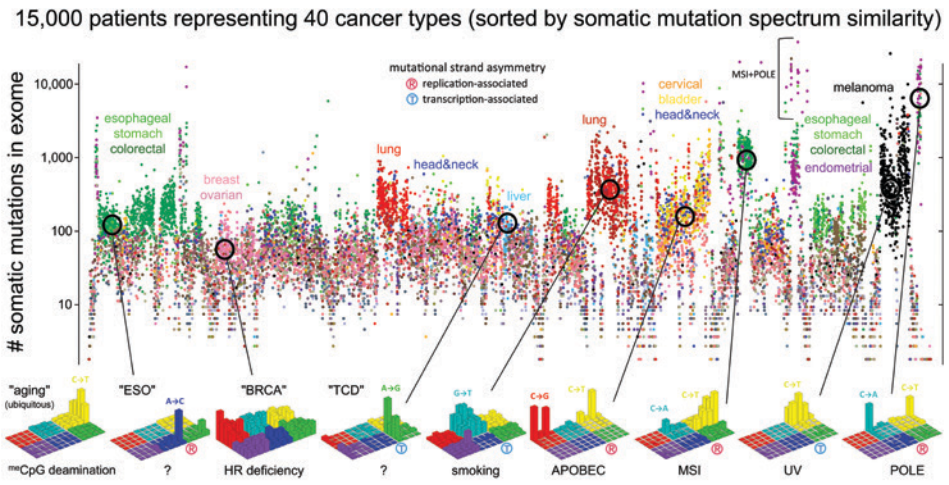
*Associate Computational Biologist based at The Broad Institute

Tumor DNA Sequencing

High-throughput DNA sequencing is a workhorse of biomedical research. There are many challenges in processing the raw DNA sequencing reads from a patient's resected tumor or biopsy material, aligning them accurately to the reference human genome, and then scanning for loci where the tumor DNA differs from the patient's bulk "normal" DNA (e.g. from a blood draw). Distinguishing true somatic mutations from sequencing or alignment artifacts can be tricky, especially for subclonal events present in only a fraction of tumor cells. We are refining a "panel of normals" (PoN) approach, which combats stochastic artifacts seen in the patient's tumor sample, and not in the patient's normal sample but widespread however in many other patients' normal samples. We are continually discovering new artifact modes, making this a highly challenging and unpredictable area of research. Isolating true somatic mutations is crucial for downstream analyses of mutational signatures and driver events.

Analyzing Mutational Signatures

Cancers vary over many orders of magnitude in their total background mutation burden, ranging from very quiet tumor types such as leukemias and childhood tumors, which may have fewer than 10 somatic mutations in their exome, to carcinogen-associated tumor types such as lung cancer and melanoma, which may have over 1000. Mutations have many causes, and each mutagen can leave a telltale signature. For instance, spontaneous deamination of methylated CpG's causes the transition mutations that dominate many tumor types. Mutagens in tobacco smoke cause G-to-T transversions. Ultraviolet radiation causes C-to-T at dipyrimidines. Agitated APOBEC enzymes cause mutations at C's preceded by T. Loss of mismatch repair causes microsatellite instability (MSI), marked by expansion and contraction of simple-



Survey of total mutational burden and mutational spectra across many tumor types. Each dot represents a cancer patient whose tumor was subjected to whole-exome DNA sequencing. Vertical position indicates the total number of somatic coding mutations in each patient. Patients are sorted by the similarity of their somatic mutation spectra (i.e. clustering dendrogram order). Colors indicate tissue types (red=lung, orange=cervical, etc.). Lower panel shows nine common mutation spectra with lines drawn to representative sample cohorts. These "Lego plots" indicate relative frequencies of the 96 possible trinucleotide-based mutation types.

sequence repeats, as well as characteristic types of single-base changes. Tumors carrying mutations in the proofreading exonuclease domain of polymerase epsilon (POLE) tend to accrue C-to-A mutations at the trinucleotide TCT. Very rare "MSI+POLE" cancers show the highest yet known somatic mutation burdens, with upwards of 10,000 coding mutations per patient. Patients affected by MSI and/or POLE mutagenesis are known to experience better clinical outcomes, probably thanks to their high neoantigen loads which attract a powerful immune response. Our most recent research has focused on a less well-studied signal in somatic mutation datasets, mutational asymmetries between the two DNA strands. These illuminate transcriptional or "T-class" mutational patterns, associated with exposure to tobacco smoke, UV radiation, and a yet-unknown agent in liver cancer, as well as replicative or "R-class" patterns, associated with MSI, APOBEC, POLE, and a yet-unknown agent in esophageal cancer.

Identifying Significantly Mutated Genes

We identify novel cancer driver genes using our algorithm MutSig (Mutation Significance). Early versions of MutSig revealed novel cancer driver genes such as DIS3 and FAM46C in multiple myeloma and SF3B1 in chronic lymphocytic leukemia. Turning to cancer types with higher levels of background mutation, such as lung cancer and melanoma, forced recognition that mutation density varies dramatically across the genome, with lowest mutation rates in regions of the genome having highest transcriptional activity, earliest DNA replication times, and most accessible chromatin structure. Incorporating these observations into MutSig enabled analysis of much larger and noisier sets of mutational data, leading to the identification of dozens of new cancer driver genes. In collaboration with wet-lab colleagues, we have begun experimentally validating these predictions in model systems.

Selected Publications:

Buisson R, **Lawrence MS**, Benes C, Zou L. APOBEC3A and APOBEC3B activities render cancer cells susceptible to ATR inhibition. *Cancer Res*. 2017 Jul 11.

Haradhvala NJ, Polak P, Stojanov P, Covington KR, Shinbrot E, Hess JM, Rheinbay E, Kim J, Maruvka YE, Braunstein LZ, Kamburov A, Hanawalt PC, Wheeler DA, Koren A, **Lawrence MS***, Getz G*. Mutational Strand Asymmetries in Cancer Genomes Reveal Mechanisms of DNA Damage and Repair. *Cell*. 2016 Jan 28; 164(3):538-49.

Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. 2014 Jan 23; 505(7484):495-501.

Roberts SA, **Lawrence MS**, Klimczak LJ, Grimm SA, Fargo D, Stojanov P, Kiezun A, Kryukov GV, Carter SL, Saksena G, Harris S, Shah RR, Resnick MA, Getz G, Gordenin DA. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet*. 2013 Sep; 45(9):970-6.

Lawrence MS*, Stojanov P*, Polak P*, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortés ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak AM, Lohr J, Landau DA, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll SA, Mora J, Lee RS, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CW, Biegel JA, Stegmaier K, Bass AJ, Garraway LA, Meyerson M, Golub TR, Gordenin DA, Sunyaev S, Lander ES, Getz G. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013 Jul 11; 499(7457):214-8.

*Co-authors