The Elizabeth H. and James S. McDonnell III
McDonnell Genome Institute
at Washington University

Washington University in St. Louis
School of Medicine

PMBIO Module 05

Somatic. Somatic WGS and Exome Variant Analysis

Malachi Griffith, Obi Griffith, Zachary Skidmore, Huiming Xia
Introduction to bioinformatics for DNA and RNA sequence analysis (IBDR01)
29 October – 2 November, 2018
Glasgow

Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)

This is a human-readable summary of (and not a substitute for) the license. Disclaimer.

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

ShareAlike — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

Learning objectives of module 05: Somatic

- **Key concepts**: Somatic variation, variant types (SNVs, small indels, CNVs, SVs, LOH), VCFs, variant allele fraction (VAF), purity estimation and tumor clonality

- Compare and contrast the merits of exome and WGS data for variant calling
- Compare germline and somatic variant calling strategies
- Consider the features of major variant types detected in NGS data
- Perform somatic variant calling of various types using tools specific to each type
- Understand the basic features of the VCF format
- Merge multiple VCFs into a single combined VCF
- Perform variant filtering to identify a high quality set of variants
- Annotate variants with respect to transcript annotations, population frequency, predicted function, etc.
- Manually review variants of each type to better understand how variant callers use read alignment data to identify variants