
LOHGIC User Guide

<http://www.khiabanian-lab.org>

October 20, 2017

Introduction

Mutations in familial cancer genes can be detected by high throughput sequencing; however, due to often lack of patient-matched control DNA and/or low tumor purity, there is limited ability to determine the genomic status of these alterations. In particular, the identification of these variants in tumor-only data raises two important questions with significant clinical implications for patient care: I) is the variant acquired somatically in the tumor, or is it in the germline, and II) is loss of heterozygosity affecting the gene's loci suggesting a role for targeted therapy in some case? The latter question, specifically, cannot be answered by routine germline testing.

LOH-Germline Inference Calculator (LOHGIC) is developed on a model-selection scheme using Akaike Information Criterion (AIC) weighting. LOHGIC infers the most consistent model describing the germline-versus-somatic mutational status, and predicts LOH for mutations identified via clinical grade, high-depth, hybrid-capture tumor-only sequencing. It also incorporates statistical uncertainties inherent to HTS as well as biases in tumor purity estimates.

LOHGIC workflow

LOHGIC [1] requires the specimen's purity and a variant's variant allele frequency (VAF), sequencing depth, and ploidy as input parameters (Figure 1). For a ploidy of Y , there are Y somatic models and Y germline models to be assessed.

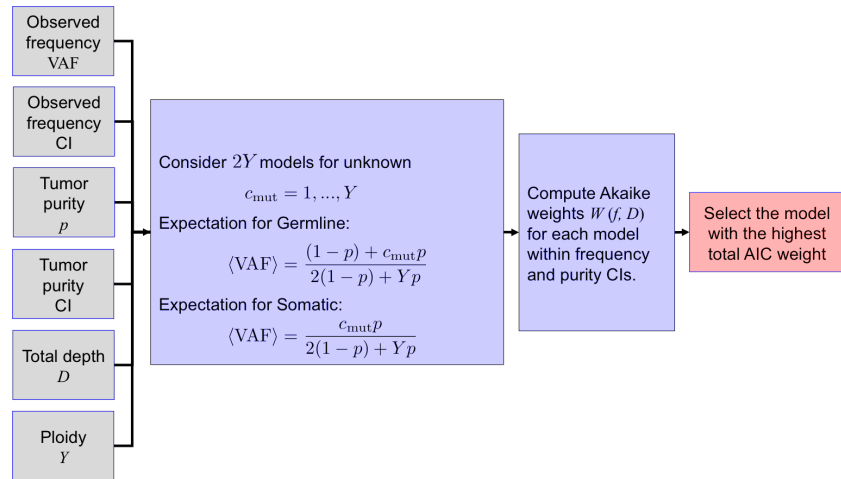


Figure 1: LOHGIC's workflow.

When ploidy is set at 2, LOHGIC also considers additional models for LOH with the loss of wild-type copy, when $Y = 1$ (Figure 2).

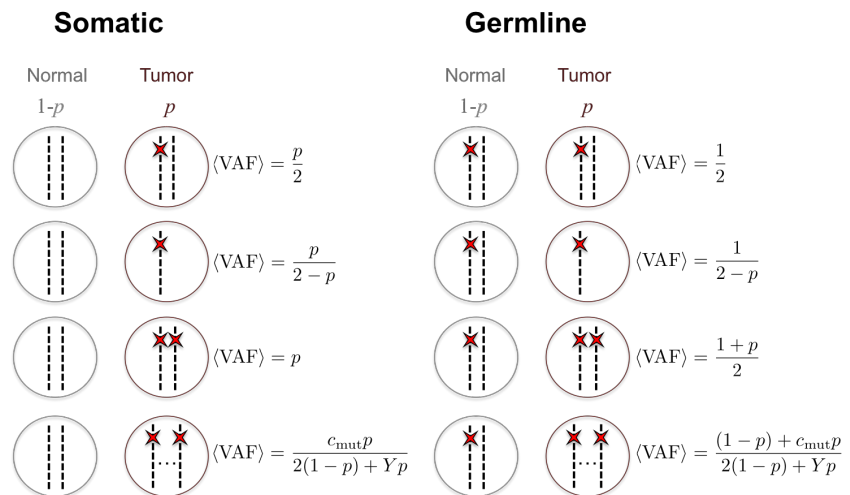


Figure 2: LOHGIC assesses somatic and germline models.

Depth of sequencing determines confidence interval (CI) for a VAF. At higher depths, CI is smaller around the reported VAF. Conversely, low sequencing depths provide limited discriminatory power to select between mutational models due to larger CIs. For example, for a variant with VAF of 0.58, sequenced at depth $950\times$, the 99% CI for VAF is between 0.54 and 0.62.

Specimen biases in tumor purity estimates also confound inferring mutational status of a variant. At very low or very high tumor purity, some combinations of ploidy and sequencing depth may not permit the selection of a model out of several, resulting in ambiguous results.

LOHGIC calculates AIC weights (W) [2, 3] within the CI for purity and VAF, and visualizes W for all possible mutational models for a give ploidy. In LOHGIC's top graph (Figure 3), the x axis depicts the CI for VAF and the shaded areas represent the CI for purity.

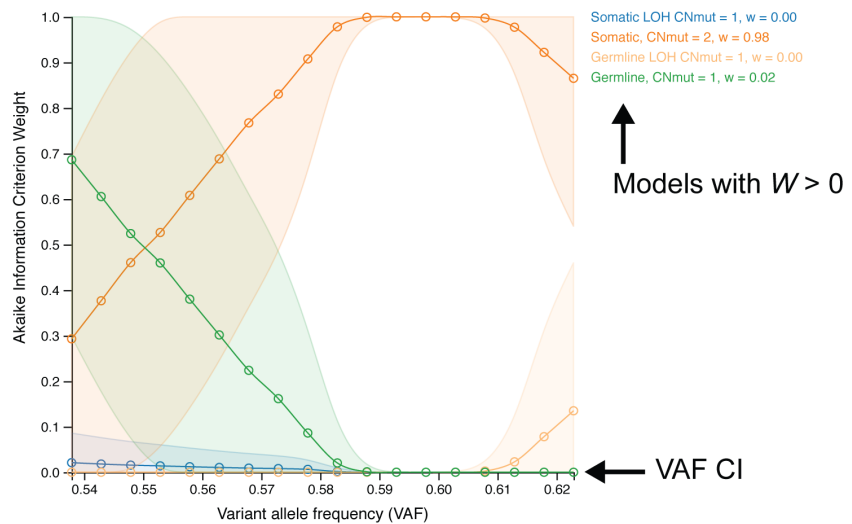


Figure 3: LOHGIC's first graph visualizes W s within the CIs.

LOHGIC also visualized the relationship between the expected VAF versus purity for each model, where the shaded gray area represents VAF and purity CIs, within which W s are calculated (Figure 4).

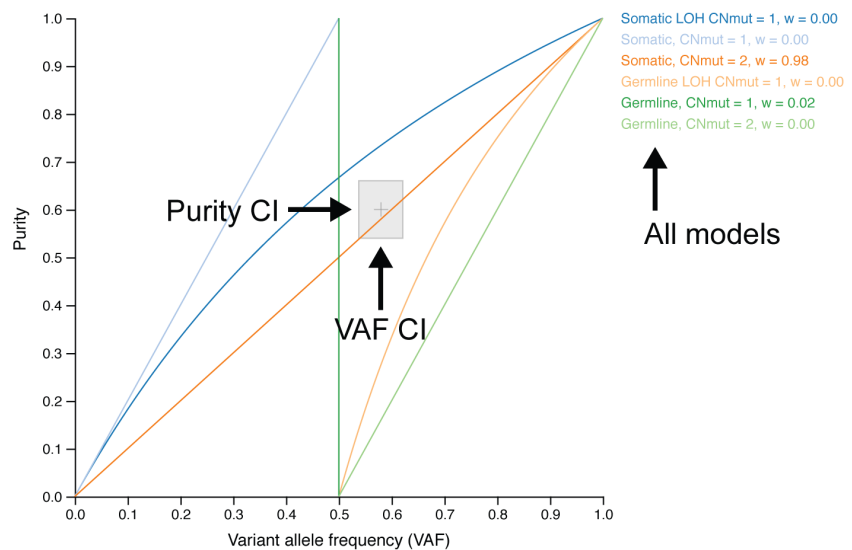


Figure 4: LOHGIC's second graph visualizes $\langle \text{VAF} \rangle$ vs. purity for all models.

The ratio of W s or their sums for a set of models indicates their relative strength of evidence. Simulations show that when mutational ploidy is less than 4, the lower bound W for unambiguous inference calls is 0.7. At higher ploidy users may need to select other appropriate thresholds.

LOHGIC is available as web-based application, implemented in JavaScript and its statistics is based on approximations implemented in MATLAB [4]. LOHGIC is also available in MATLAB upon request.

References

- [1] H. Khiabani, K. M. Hirshfield, M. Goldfinger, S. Bird, M. Stein, J. Aisner, D. Toppmeyer, S. Wong, N. Chan, K. Dhar, J. Gheeya, H. Vig, M. Hadigol, D. Pavlick, S. Ansari, S. Ali, B. Xia, L. Rodriguez-Rodriguez, and S. Ganesan, "Inference of germline mutational status and evaluation of loss of heterozygosity in high-depth tumor-only sequencing data," *JCO Precision Oncology*, vol. In press, 2017.
- [2] H. Bozdogan, "Model selection and akaike's information criterion (aic): The general theory and its analytical extensions," *Psychometrika*, vol. 52, pp. 345–370, Sep 1987.
- [3] E.-J. Wagenmakers and S. Farrell, "Aic model selection using akaike weights," *Psychonomic Bulletin & Review*, vol. 11, pp. 192–196, Feb 2004.
- [4] C. Loader, "Fast and accurate calculations of binomial probabilities." July 2000.