# Double Sequencing Analysis (this still needs a title)

Dakota Z. Derryberry, Claus O. Wilke, Matthew C. Cowperthwaite

May 23, 2014

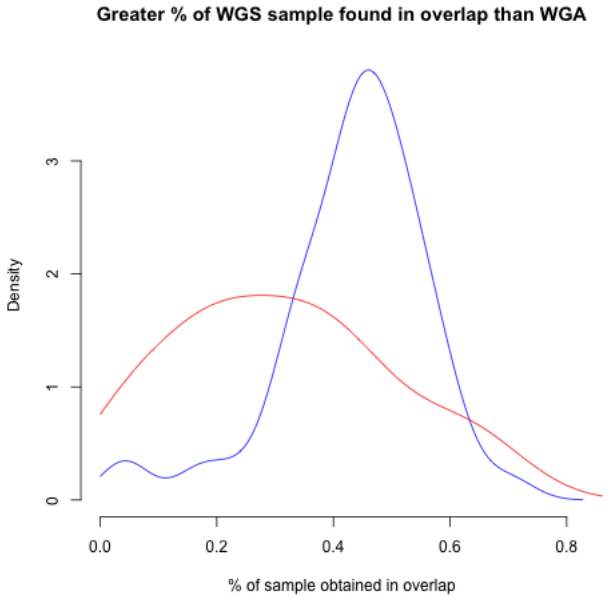
## 1 Introduction

Intro goes here.

## 2 Results

### 2.1 Unfiltered data: WGA v. WGS

As expected, we found that



### 2.2 Unfiltered data: Time points

### 2.3 Filtering

## 3 Methods

### 3.1 Data and back-end processing

All sequence data came from The Cancer Genome Atlas (TCGA) Gliobastoma multiforme (GBM) data set. We downloaded raw reads in (XXX format) using (CGHub?) on (date) for 68 patients. For each patient, data consisted of one set of rads taken from blood DNA, and two sets of reads taken from tumor DNA. In 55 cases, the two sets of reads from tumor DNA were one set of reads from whole genome sequencing (WGS) and one set of reads from whole genome sequencing with amplification (WGA). In 13 cases, the two sets of reads from tumor DNA were one set of reads pre-readiation treatment and one set post-radiation treatment. We developed a pipeline to align all reads to HG19. This pipeline used (bowtie?) for alignment and (bedtools? whatever – quality control...) We used SomaticSniper to align each tumor sequence with it's corresponding blood sequence. We then filtered the SomaticSniper data using 8 custom filters. These removed (i) , (ii) , (iii) , (iv) , (v) , (vi) , (vii) , and (viii) . All custom code is available in a github repository, located at (webaddress). In addition to sequence data, filter 6 uses dbSNP, version 137, to find common SNPs among the putative mutaitons called by SomaticSniper.

### 3.2 Analysis

We used custom python scrips, available in the above github repository, to perform simple calculations and data operations. We used R to do statistics and generate figures, and this code is also availabel in the git repository.

## 4 Discussion

What do I think of this?