Question 1

Report

It is evident from the article that the several observations in regard to the experiment are not deviant with initially approved researches on tumor purity. We initially examined the relationships between the number of mutations detected by four mutation calling techniques as well as the associated tumor purities using gastric cancer as a model. Purities of stomach cancer tumors varied widely, ranging from 5% to 100%. Around 72% of gastric tumors had purity greater than 70%.

The results indicated a substantial positive correlation between the number of mutations detected by MuSE and SomaticSniper algorithms and tumor purity ($p = 1.21e-05$ for MuSE and $p = 1.64e-04$ for SomaticSniper). Similar results were reported for the number of mutations detected by the MuTect2 ($p = 1.01e-02$) and VarScan2 ($p = 6.13e-05$) algorithms, which incorporate tumor purity correction criteria. Notably, the considerably positive connection between the number of mutations and tumor purity was also detected in the other nine cancer types. These findings suggested that tumor purities may have a major effect on mutation discovery.

It is critical to note that when normal samples from multiple tissue types are combined, the diversity within the normal group may rise. This is why, in an earlier version of InfiniumPurify, we identified iDMCs using matched samples in each cancer type. However, rigorous data analysis reveals that mixing normal samples provides outcomes that are equivalent. This, we feel, is a more important tactic that will have a broader application; for instance, purity prediction can be conducted for malignancies not included in the TCGA. For every form of cancer, as long as even the sample size is sufficiently high (e.g. 20,) the iDMCs can be

consistently recognized and the purity determined by correlating the cancer to universal normal controls.

It is clear that tumor purity estimations display inherent features of the source data utilized to assess purity and have only a moderate correlation with other profiles. One possible explanation for these differences is the beginning tissue material used to examine the various locations of the tumor specimen. Estimates based on pathology are considered the top standard.

However, the interpathologist variation seen in this study, as well as previous research, implies that these estimations are likely to contain some mistakes due to their subjectivity. These inconsistencies may possibly be due to the pathologic slide's lack of complete spatial heterogeneity. For clinico-genomic sequencing investigations that require a certain level of purity for inclusion, an alternative to pathology estimations is to deduce purity explicitly from the analyte by doing moderate DNA sequencing to filter out low-purity samples.

Additionally, emphasis is on DM calling methods that require a larger sample size than minfi, limma, or equivalent tools. It is anticipated that InfiniumPurify would be used mostly for population-level research. Control-free DM calling additionally demands that the purities of the samples be sufficiently distributed to allow the statistical test to be done correctly. It is vital using the regulation DM calling method with prudence and using normal controls whenever possible.