TCGA ASE Analysis Pipeline

```
0_Download_SNPArray_Table_From_GDC.pl
```
↓ ●Get SNPArray IDs for downloading data

```
0_Download_SNPArray_From_GDC.pl
```
→ `1.0_Prep_SNPs_for_Imputation.pl`

●Downloads SNPArray data from GDC

●Prepares affy6 files for imputation

```
1.1_Birdseed_to_ped_and_maps.pl
```

●Makes ped and map files from downloaded SNPArray data

1000 Genomes Project
```
1092 Whole-Genome phased haplotypes
```
→ `1.2_Shapeit_and_Imputation.pl`

●Runs shapeit and imputation along with the haplotypes from the genome project

```
1.3_Make_het_cds.pl
```

●Keeps het snps and makes cds sorted files

```
2.0_Lookups_for normal.pl
```

```
3.0_Dwnld_Table_4_RNASeq_From_GDC.pl
```

●Lookups the cds files and gets the normal samples

●Get RNA-Seq or WGS IDs for downloading data

```
2.1_Get_Bad_SNPs_and_CNVs.pl
```

●Get bad SNPs and CNVs of the cancer type

```
3.0_Download_RNA_Seq_WGS_and_do_Mpileup.pl
```

●Downloads RNA-Seq or WGS data and runs mpileups

RNA-Seq   WGS

```
3.1_ASE_Analysis.pl
```
```
4.0_Somatic_Variants.pl
```

●Performs gene level ASE analysis

●Uses varscan to get somatic variants

```
3.2_Export_ASE_Data.pl
```·······`4.1_Upstream_Downlstream_Analysis.pl`

●Get gene-level reads by summarizing phased het variants and gets it ready to be analyzed

●Performs upstream and downstream analysis and gets data ready to be analyzed

```
Correlation analyses between allelic imbalance genes and the occurence of mutations/indels
```

●Map these mutations/indels to 18 annotation regions, including exons, 3UTR, 5UTR, and so on

●Obtain top allelic imbalanced genes associated with mutations /indels mapped to 18 different annoation regions