To annotate large number of available NGS data, there is a large number of SNPs annotation tools available. Some of them are specific to some specific SNPs annotation. Some of the available SNPs annotation tools are: SNPeff, VEP, ANNOVAR, FATHMM, PhD-SNP, PolyPhen-2, SuSPect, F-SNP, AnnTools, SeattleSeq, SNPit, SCAN, Snap, SNPs&GO, LS-SNP, Snat, TREAT, TRAMS, Maviant, MutationTaster, SNPdat, Snpranker, NGS – SNP, SVA, VARIANT, SIFT, PhD-SNP and FAST-SNP.

We decided to work with ANNOVAR. This tool is suitable for pinpointing a small subset of functionally important variants. Uses mutation prediction approach for annotation. An important and probably highly desirable feature is that ANNOVAR can help identify subsets of variants based on comparison to other variant databases, for example, variants annotated in dbSNP or variants annotated in 1000 Genome Project. The exact variant, with same start and end positions, and with same observed alleles, will be identified.

Variant calling, made by the Gerstein Lab located 3,559,138 SNPs and 789,969 indels. Using ANNOVAR, we performed a comparison against 1000genome (version 1000g2015aug) and gnomAD (both in build hg19).

A comparison of subjectZ with all individuals in 100gemone: 95% of the SNPs (~3.4M) and 60% of the indels (469K) were found within some other people.
A comparison of subjectZ with all individuals in gnomAD: 97.5% of the SNPs (~3.5M) and 90% of the indels (709K) were found within some other people.

Source	Known variants (dropped)	Unknown variants (filtered)	Known variants (dropped)	Unknown variants (filtered)	Total
	1000Genome		gnomAD		
indel	469,754	320,215	709,146	80,823	789,969
snp	3,381,358	177,780	3,476,922	82,216	3,559,138

As we can see above, there is a variance of ~2.5-30% when comparing to 1000genome and gnomAD. gnomAD find much more variants, as we expect, since it has a larger individual cohort.

Appendix – sample command lines:

Download relevant databases:

1. annotate_variation.pl -downdb 1000g2015aug humandb -buildver hg19
2. annotate_variation.pl -buildver hg19 -downdb -webfrom annovar gnomad_genome humandb/

VCF conversion:

1. convert2annovar.pl -format vcf4 -withfreq data/indel.vcf > data/indel.avinput
2. convert2annovar.pl -format vcf4 -withfreq data/snp.vcf > data/snp.avinput

Analysis:

1. annotate_variation.pl -filter -dbtype 1000g2015aug_all -buildver hg19 -out indel data/indel.avinput humandb/
2. annotate_variation.pl -filter -dbtype hg19_gnomad_genome -buildver hg19 -out snp.gad data/snp.avinput humandb/

Sources:

1. Wikipedia web site - SNP annotation
2. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data, Nucleic Acids Research, 38:e164, 2010