Claims

1. A method of annotating a variant nucleic acid sequence in a sequence database, comprising the steps:
    a) carrying out high-throughput sequencing of at least one nucleic acid sequence of interest to obtain multiple sequence reads across the sequence of interest;
    b) analysing said sequence reads to determine variants in the sequence of interest that co-occur;
    c) grouping sequence reads with said co-occurring variants;
    d) comparing grouped co-occurring variants to a sequence database comprising one or more reference sequences, wherein the database comprises at least one wild-type sequence corresponding to the nucleic acid sequence of interest; and
    e) thereby annotating the sequence in the sequence database corresponding to the nucleic acid sequence of interest to include information relating to the co-occurring variants.

2. The method according to claim 1 wherein the database further comprises at least one variant of the wild-type sequence.

3. The method according to claim 1 or 2, wherein the database is the COSMIC database.

4. The method according to any one of claims 1 to 3, further comprising determining the genomic DNA sequence and coding DNA sequence corresponding to an annotated variant.

5. The method according to any one of claims 1 to 4, further comprising a step, prior to step a) or b), of filtering out variants that fall outside a region of interest in the nucleic acid sequence of interest.

6. The method according to claim 4, further comprising determining the encoded amino acid sequence corresponding to the coding DNA sequence.

7. The method according to any one of claims 1 to 6, wherein the variant is selected from a single nucleotide polymorphism (SNP), multiple nucleotide polymorphism, a mutation within codons spanning an intron, an insertion, and a deletion.

8. The method according to any one of claims 1 to 7, wherein the variant encodes an amino acid sequence comprising at least one altered amino acid residue compared with a wild type amino acid sequence.

9. The method according to claim 8, wherein the at least one altered amino acid residue results in an altered susceptibility or responsiveness of the encoded polypeptide or protein to a drug.

10. The method according to claim 9, wherein the drug is an anti-cancer drug, anti-viral drug, or anti-bacterial drug.