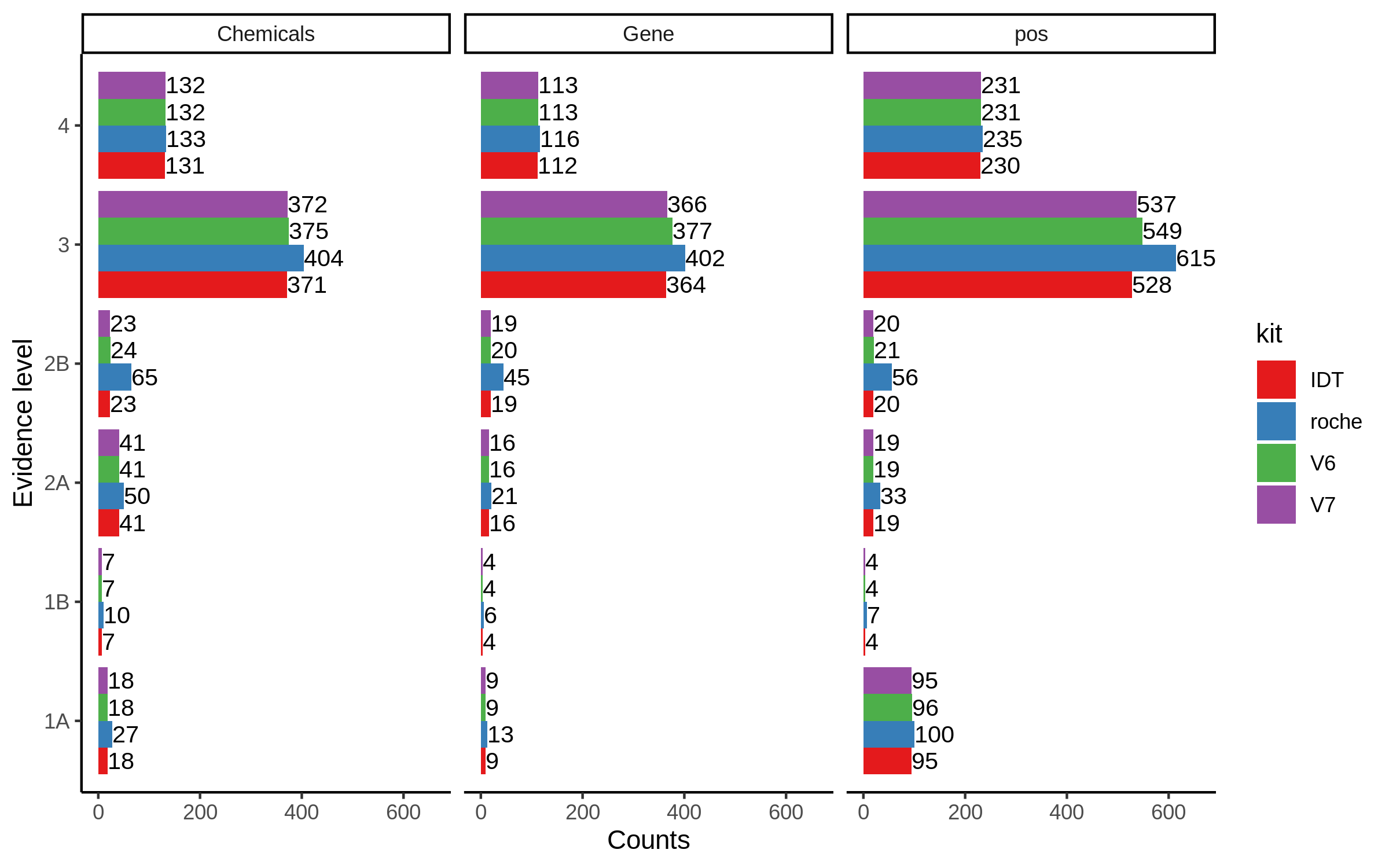
1. 請將 test.vcf.gz 以 Transcript 為單位轉成 TSV 格式(tab 分隔)
   1. 備註
      1. VCF 格式中，會以 ; 作為註解的分隔號
      2. CSQ 註解中會以 , 作為 Transcript 的分隔號
      3. CSQ 註解中會以 | 作為 每個欄位的分隔號
      4. CSQ 欄位的標頭可從 VCF header 中，CSQ 描述中取得
      5. 略過沒有 CSQ 註解的部分
   2. 輸出欄位
      1. CHROM
      2. POS
      3. REF
      4. ALT
      5. FILTER
      6. Consequence
      7. SYMBOL
      8. Gene
      9. Feature
      10. HGVSc
      11. HGVSp
2. 請根據 reference.csv.gz 的 probe 資訊，計算 input.tped 中，每個 probe 的統計結果，輸出成 TXT
   1. 備註
      1. reference.csv.gz 中，Probe Set ID 欄位代表 Probe\_id
      2. reference.csv.gz 中，Alt Allele 有多種可能時，會以 "/" 當分隔號表示
      3. input.tped 格式中，第2欄為 probe\_id，第5欄開始為 sample call
      4. input.tped 中，sample call 如果為 "0 0" 代表 No-call
   2. 輸出欄位
      1. probe\_id
         1. 轉大寫
      2. call\_rate
         1. 不是 No call 的樣本數 佔多少
      3. allele\_count
         1. 不是 No call 的樣本數
      4. allele\_frequency
         1. alt allele 佔多少 %
      5. heterogeneity
         1. alt allele == 1 的有多少
      6. homogeneity
         1. alt allele == 2 的有多少
   3. 請將上一步的程式寫成dockerfile, 輸入的變數為reference.csv.gz, input.tped, 輸出的變數則為TXT檔, 輸入及輸出需要以掛載的方式進行
3. 請從A\_gen.fasta檔案中分別輸出HLA:HLA25681 HLA:HLA05443 HLA:HLA26560 HLA:HLA26769 HLA:HLA26076中第251-270之間的序列   
   並請以任意工具以柱狀圖比較5條序列中ACGT的個數
4. 附檔中的bed檔為library prep kit的bed檔(hg38), 請找出clinical\_ann\_metadata.tsv 欄位Location為rsid的位點, 是否為在bed檔的含蓋範圍, 並以Level of Evidence 分別繪出各個bed檔含蓋的位點數(POS), 基因數(Gene), 藥物數(Related chamicals), 如下



1. 請閱讀附檔中的文獻並回答以下問題
   1. 這篇文獻主要是在評比什麼樣的工具, 為什麼會需要用到這類工具.
   2. 這篇文獻評比的結論是什麼, 為什麼評比結果較好的工具能夠有比較好的表現?
   3. 如果我們要做一個類似的工具, 會需要用到哪些素材來建立一個類似的工具
2. 附檔中的樣本患某種遺傳疾病, 目前已知相關的表徵為Nystagmus, Visual impairment, Global developmental delay, Optic atrophy,Retinitis pigmentosa inversa.
   1. 請從位點清單中, 找出可能的致病位點, 及推斷出其疾病名稱
   2. 或是描述出可能可以找到致病位點或疾病的方法