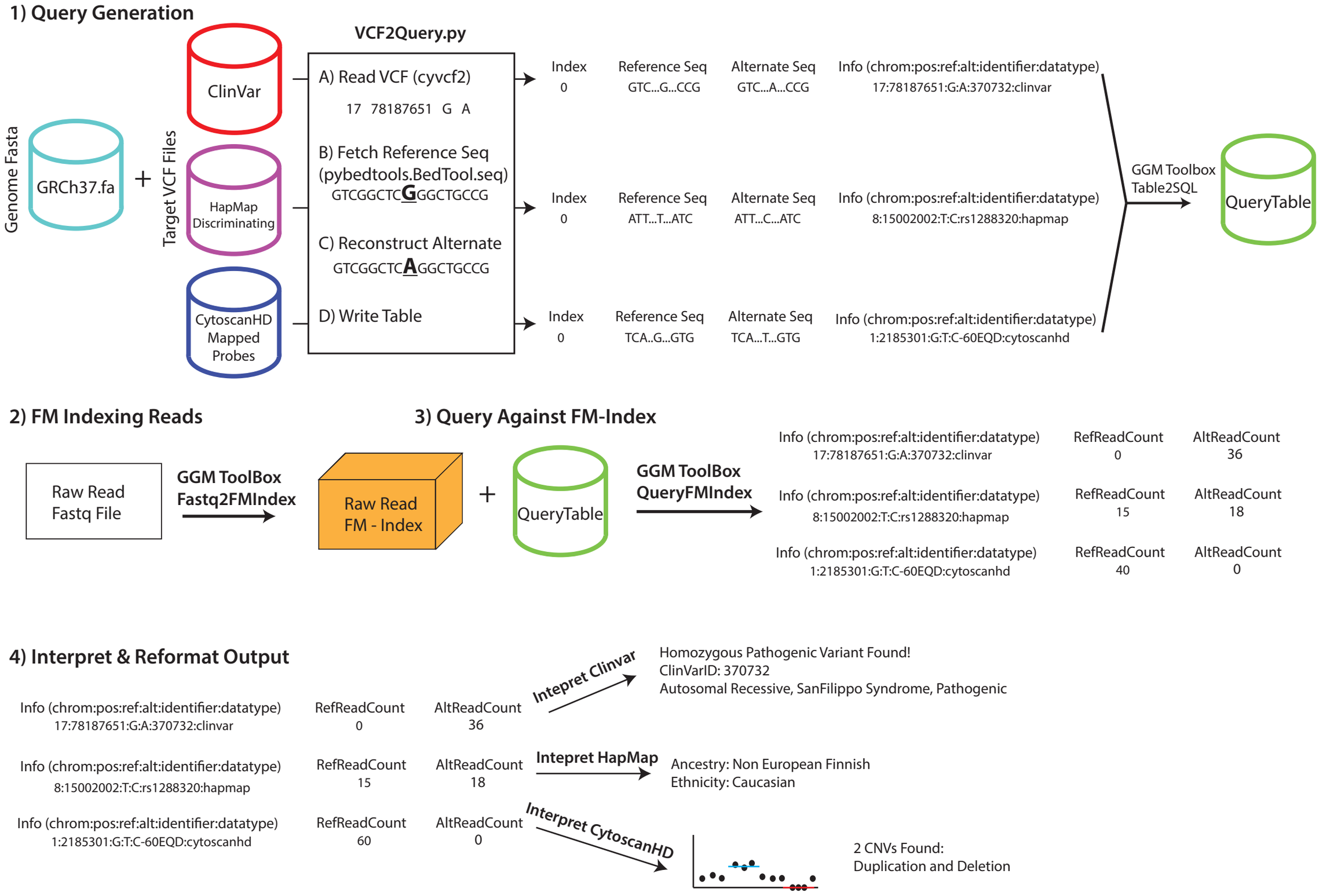
1) Query Generation
VCF2Query.py
Genome Fasta
GRCh37.fa
Target VCF Files
ClinVar
HapMap Discriminating
CytoscanHD Mapped Probes
A) Read VCF (cyvcf2)
17 78187651 G A
B) Fetch Reference Seq (pybedtools.BedTool.seq)
GTCGGCTCGGGCTGCCG
C) Reconstruct Alternate
GTCGGCTCAGGCTGCCG
D) Write Table
Index
Reference Seq
Alternate Seq
Info (chrom:pos:ref:alt:identifier:datatype)
0
GTC...G...CCG
GTC...A...CCG
17:78187651:G:A:370732:clinvar
0
ATT...T...ATC
ATT...C...ATC
8:15002002:T:C:rs1288320:hapmap
0
TCA..G...GTG
TCA...T...GTG
1:2185301:G:T:C-60EQD:cytoscanhd
GGM Toolbox
Table2SQL
QueryTable
2) FM Indexing Reads
Raw Read
Fastq File
GGM ToolBox
Fastq2FMIndex
Raw Read
FM - Index
3) Query Against FM-Index
QueryTable
GGM ToolBox
QueryFMIndex
Info (chrom:pos:ref:alt:identifier:datatype)
RefReadCount
AltReadCount
17:78187651:G:A:370732:clinvar
0
36
8:15002002:T:C:rs1288320:hapmap
15
18
1:2185301:G:T:C-60EQD:cytoscanhd
40
0
4) Interpret & Reformat Output
Info (chrom:pos:ref:alt:identifier:datatype)
RefReadCount
AltReadCount
17:78187651:G:A:370732:clinvar
0
36
Intepret Clinvar
Homozygous Pathogenic Variant Found!
ClinVarID: 370732
Autosomal Recessive, SanFilippo Syndrome, Pathogenic
Info (chrom:pos:ref:alt:identifier:datatype)
RefReadCount
AltReadCount
8:15002002:T:C:rs1288320:hapmap
15
18
Intepret HapMap
Ancestry: Non European Finnish
Ethnicity: Caucasian
Info (chrom:pos:ref:alt:identifier:datatype)
RefReadCount
AltReadCount
1:2185301:G:T:C-60EQD:cytoscanhd
60
0
Interpret CytoscanHD
2 CNVs Found:
Duplication and Deletion