**Theory**

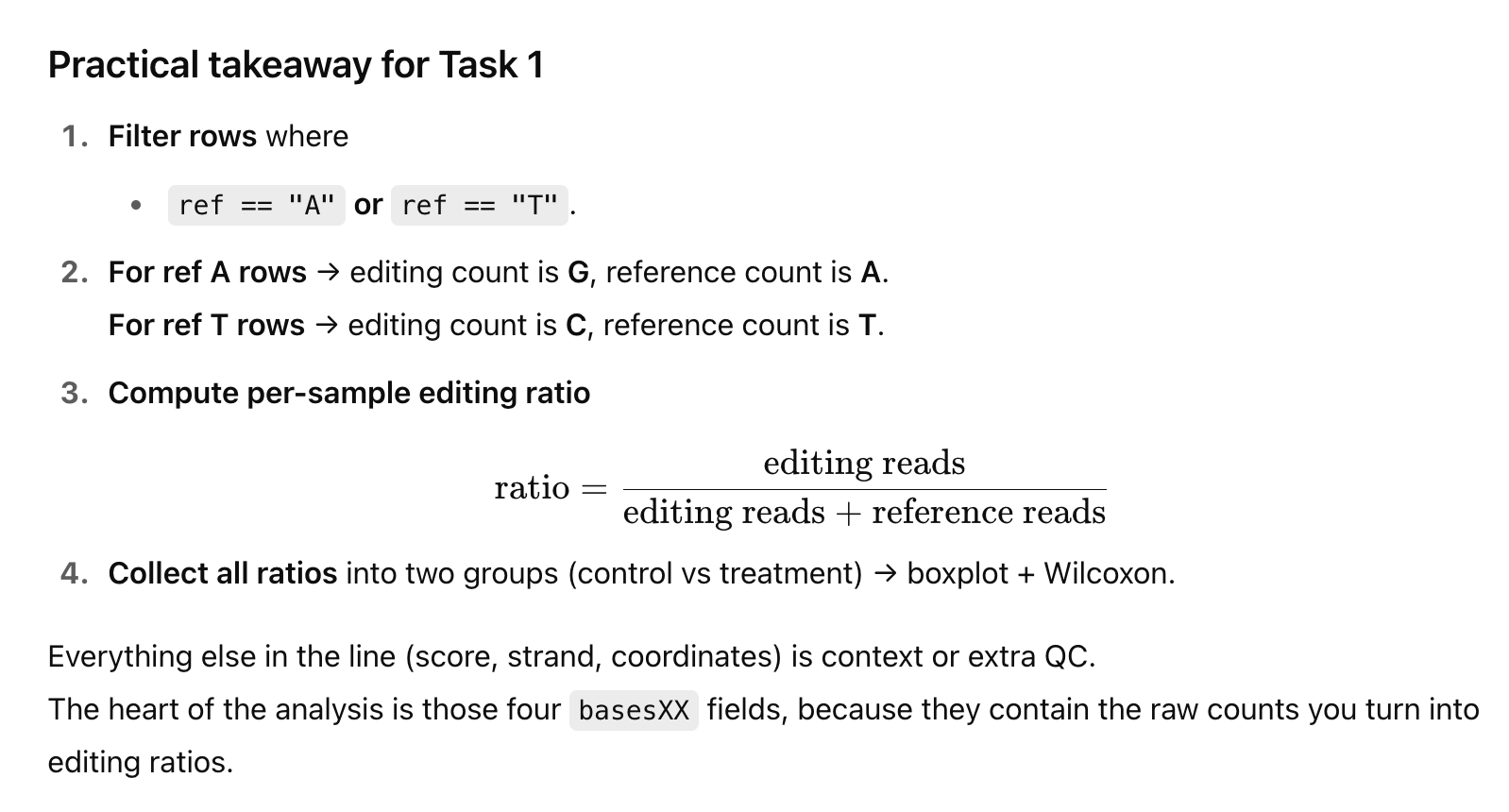
JACUSA2 ,Which compares:

**Two control BAMs  
Two treatment BAMs  
Which reference base is shown in the ref column?**

It is taken from the hg38 reference genome I aligned to (GRCh38.primary\_assembly.genome.fa).

For every 1-bp window (start–end) JACUSA2 looks up that base in the reference and writes it into the ref column

And identifies **variant-like events**, such as **A-to-G mismatches**, across the transcriptome.



**# 1. How many rows are A‐to‐I candidates?**

grep -v '^#' jacusa2\_control\_vs\_treat.tsv | \

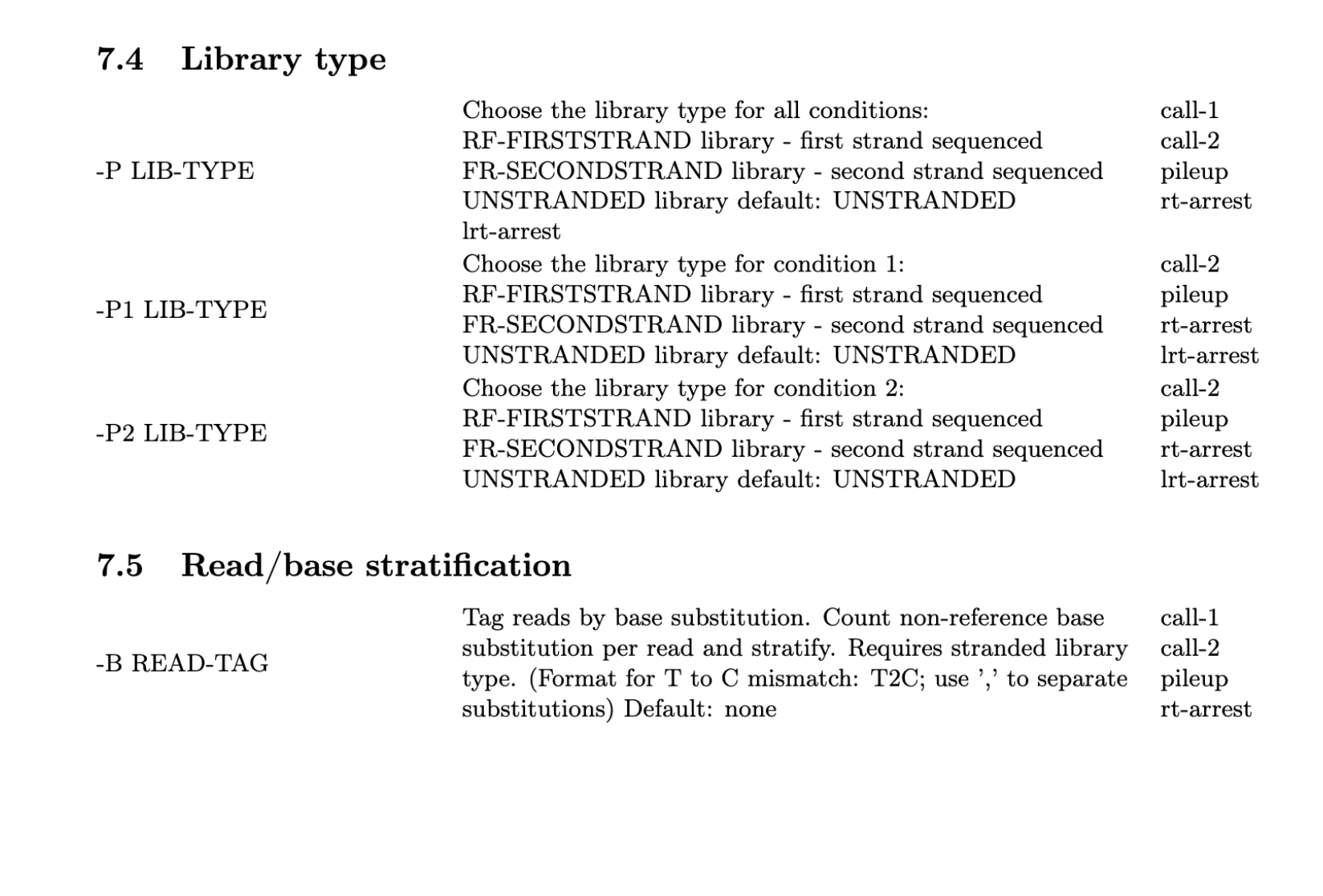
awk '$13=="A" || $13=="T"' | wc -l

**# 2. Extract BED of candidates for IGV or overlap checks**

grep -v '^#' jacusa2\_control\_vs\_treat.tsv | \

awk '$13=="A" || $13=="T"{print $1"\t"$2"\t"$3}' > editing\_sites.bed

*From the github page:https://github.com/dieterich-lab/JACUSA2/blob/master/manual/manual.pdf*



### **Why A2G,T2C really does cover both strands**

* If the **reference base is A** (plus strand), a deaminated inosine is sequenced as **G**.
* If the **reference base is T** (i.e., the read comes from the minus strand of an A), the inosine shows up as **C**.

So A2G + T2C together capture *every* genomically encoded A that was edited, regardless of strand.

| **Flag** | **What it does inside JACUSA2** | **Why it is needed for *stranded* A-to-I editing** |
| --- | --- | --- |
| call-2 | Compare **two groups** of BAMs (condition 1 vs condition 2). | Perfect fit for **Ctrl (DMSO)** vs **Treat (dTAG)**. |
| *BAM lists* | 1st comma-separated list = both control replicates; 2nd list = both treatment replicates. | JACUSA2 pools replicates **within** each condition when I ask it for per-condition counts. |
| -P RF-FIRSTSTRAND | Tells JACUSA2 the library orientation: • *R* = read 1 maps to **reverse** genomic strand• *F* = read 2 maps to forward• “FIRSTSTRAND” = first cDNA strand sequenced. | Matches Illumina TruSeq Stranded (the kit almost everyone uses). **Required** so JACUSA2 can decide which genomic base an A->G or T->C mismatch belongs to. |
| *(optional)* -P1 / -P2 | Same as -P but I can set each condition separately. | Only needed if library preps differ between control and treatment (they don’t here). |
| -B A2G,T2C | **Tag only A→G mismatches** (on whatever strand the read aligns) **and T→C mismatches** (the reverse-strand mirror). | This isolates the **two substitution classes produced by inosine**. All 10 other mismatch types are ignored, which prevents them from flooding my site list. |
| -f V | Write **VCF output**. In VCF each record has DP4= A\_fwd, A\_rev, G\_fwd, G\_rev (or T,C) counts **per condition**. | I now have *exact* A + G (or T + C) numbers to compute the editing ratio. BED-extended format only gives “matches vs substitutions” and therefore is **not** base-specific. |
| -R /path/to/GRCh38.fa | Needed only when -f V so JACUSA2 can fill the REF allele field in VCF. | Without it the VCF writer aborts. |
| -q 20 | Drop bases with Phred < 20 (≥ 1 % error rate). | Standard quality filter used in the paper. |
| -p 8 | Use 8 worker threads. | Scales linearly until disk I/O saturates. I tried 12 or 16, but the cluster didnt let me. |
|  |  |  |
|  |  |  |
|  |  |  |