Have you had a chance to watch the videos and practice at all with the UCSC Genome browser? I have a list of SNPs associated with RA that another student compiled from the literature. I would like you to find these in UCSC and download the SNP sequence, +/- 50 base pairs. The idea is to match these flanking bases against actual genomic sequence and compare which allele (reference or alternate) is present at the actual SNP position. We need the flanking bases to ensure (or maximize the chance) that we are finding that specific position in the genome. We would essentially be skipping the alignment step, which is very time consuming on standard computers.

Here’s the list of SNPs:

|  |  |  |  |
| --- | --- | --- | --- |
| **ID** | **Position** | **Gene** | **Odds ratio** |
| 6p21 | rs6910071 | 32315077 | HLA-DRB1 | 2.88 |
| 1p13 | rs2476601 | 113834946 | PTPN22 | 1.94 |
| 6q23 | rs5029937 | 137874014 | TNFAIP3 | 1.40 |
| 6q23 | rs6920220 | 137685367 | TNFAIP3 | 1.22 |
| 8p23 | rs2736340 | 11486464 | BLK | 1.21 |
| 2q32 | rs7574865 | 191099907 | STAT4 | 1.16 |
| 1p13 | rs11586238 | 116134528 | CD2,CD58 | 1.13 |
| 1q23 | rs12746613 | 161497252 | FCGR2A | 1.13 |
| 2p16 | rs13031237 | 60908994 | REL | 1.13 |
| 9q33 | rs3761847 | 120927961 | TRAF1,C5 | 1.13 |
| 2q11 | rs10865035 | 100219272 | AFF3 | 1.12 |
| 2q33 | rs1980422 | 203745673 | CD28 | 1.12 |
| 6q21 | rs548234 | 106120159 | PRDM1 | 1.10 |
| 9p13 | rs2812378 | 34710263 | CCL21 | 1.10 |
| 22q12 | rs3218253 | 37148770 | IL2RB | 1.09 |
| 10p15 | rs2104286 | 6057082 | IL2RA | 0.92 |
| 6q23 | rs10499194 | 137681500 | TNFAIP3 | 0.91 |
| 6q25 | rs394581 | 159061489 | TAGAP | 0.91 |
| 12q13 | rs1678542 | 57574932 | KIF5A,PIP4K2C | 0.91 |
| 4q27 | rs6822844 | 122588266 | IL2,IL21 | 0.90 |
| 1p36 | rs3890745 | 2622185 | TNFRSF14 | 0.89 |
| 1q31 | rs10919563 | 198731313 | PTPRC | 0.88 |
| 11p12 | rs540386 | 36503743 | TRAF6 | 0.88 |
| 2q33 | rs3087243 | 203874196 | CTLA4 | 0.87 |
| 10p15 | rs4750316 | 6351298 | PRKCQ | 0.87 |
| 20q13 | rs4810485 | 46119308 | CD40 | 0.85 |

Start at the top, if this proves too time consuming we can test with a subset. For each SNP we need to know what the alleles are for that position and again the flanking bases. Also, using the browser, look for other common SNPs (using the common SNPs track) within that flanking region. It’s possible there won’t be any, but we should check.