All of this analysis is from the pediatric control and MS patients from the following two studies: <http://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP056954>

<http://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP052860>

as well as the TCR sequences from pediatric patients prepared by Liwen and Xaiqing using 5’RACE.

**V and J gene profiles:**

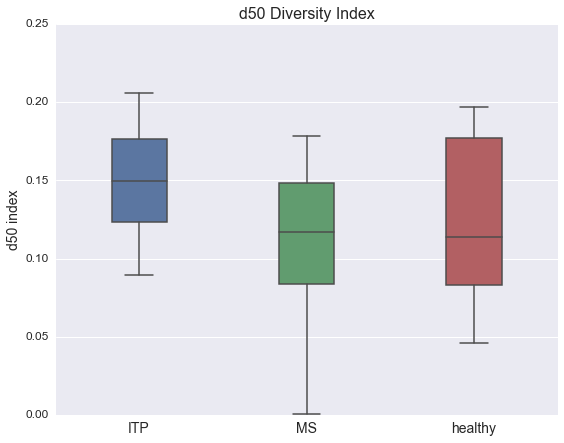
Overall, there is a strong difference between the ITP and MS data in terms of their V and J gene usage. However, this could very easily be due to different sample prep techniques – so I might not read too much into it.

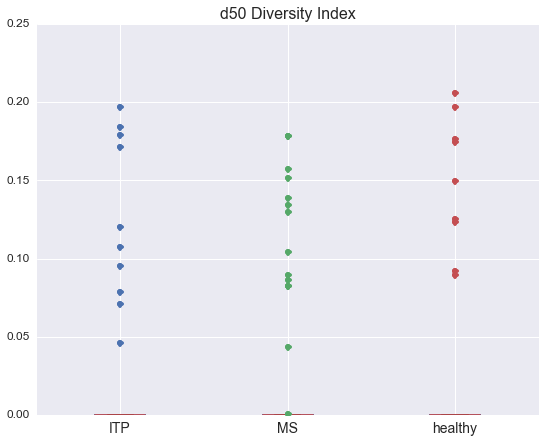
Interestingly enough, we spoke to Jacob Glanville yesterday and he mentioned that some V genes are more likely to make it into memory T-cells than others. Often times when he sees repertoires with different V gene usage, the reason is not due to different people having different proportions of memory vs naïve T-cells in the T-cell compartment. It could be interesting to explore this further, as it may inform some of the diversity analysis as well.



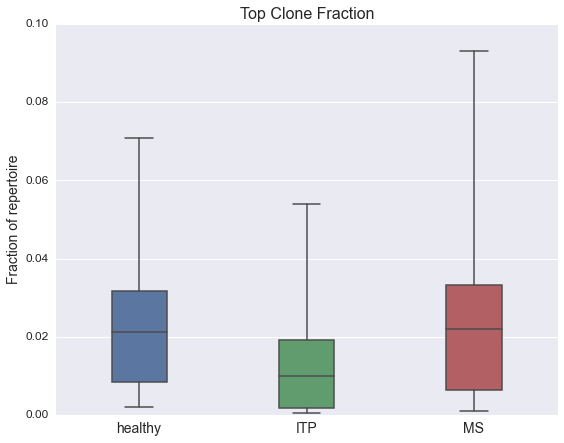
**Diversity:**

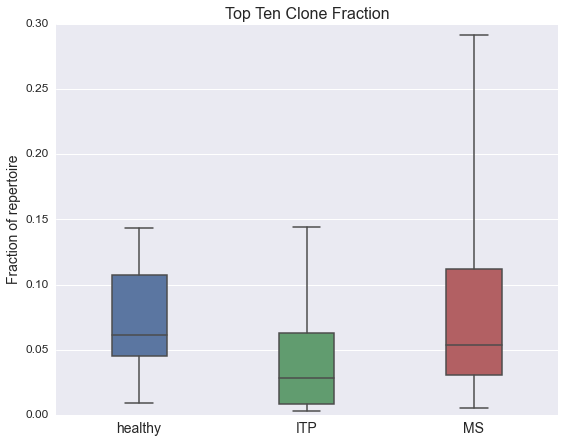
There is a couple ways I analyzed diversity. The first is by using the d50 index, which is a statistic that is calculated by determining what percentage of the total clones are needed to make up 50% of the total reads in the population. It gives a good idea of general clonal expansion. The plots below show exactly the same data, just one break its out into individual dots vs the box and whisker. A larger d50 index represents a more diverse repertoire.

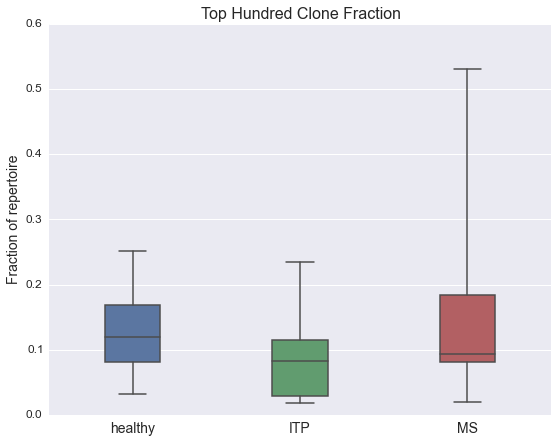




The next thing I wanted to look at was the fraction of the repertoire that the largest clone took up, the largest 10 clones, and the largest 100 clones, for each group. In each case, a smaller value represents less clonal expansion at the top of the repertoire.



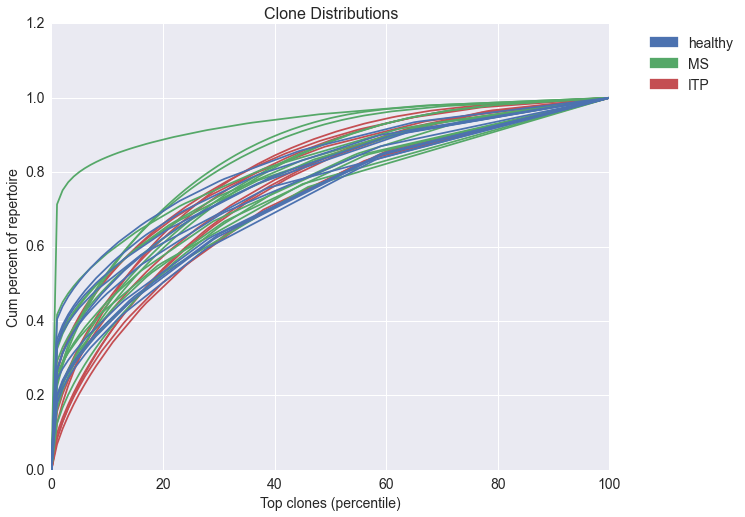


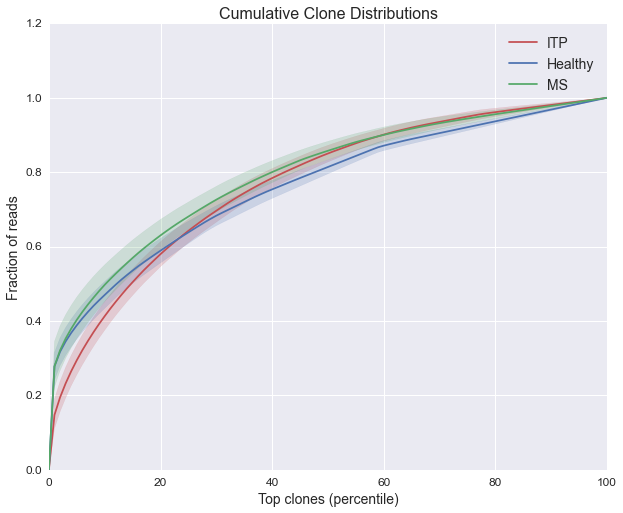


The metrics above (d50 and top clone fractions) confirm your previous analysis, that ITP in general is more ‘diverse’ or at least is less dominated by large clonal expansions. I am very excited to look at B cell repertoire data in these patients to see if the same trends continue in the B-cell compartment. In addition, with B-cells there are several other metrics (class switching, somatic hypermutation etc.) that can be used to drill further down into the mechanism behind the increase in diversity.

Finally, just to get a better sense of the overall clonal distributions, I plotted the cumulative distribution of total reads vs unique reads (‘clones’). The first plot has every patient represented by it’s own line, the second plot groups the patients together and gives the average and stdev (shaded area).

There is one MS clone that has a huge clonal expansion at the beginning (this is the patient that recently had a HSCT). Most of the other distributions are more similar. However, while healthy and MS have a more uneven distribution of clones (a few clones which are highly prevalent followed by a long tail of more rare clones) the ITP patients show a trend of having a more even distribution.





I worry a bit that the differences we see might be due to differences in prep. In particular, could the amount of starting material that went into the library creation and sequencing account for the differences we see between MS/healthy vs ITP? I’m not quite sure that is the case, but I am thinking of ways that we could test (for example, by down sampling the data so that we do analysis on the same amount of reads for each patient).