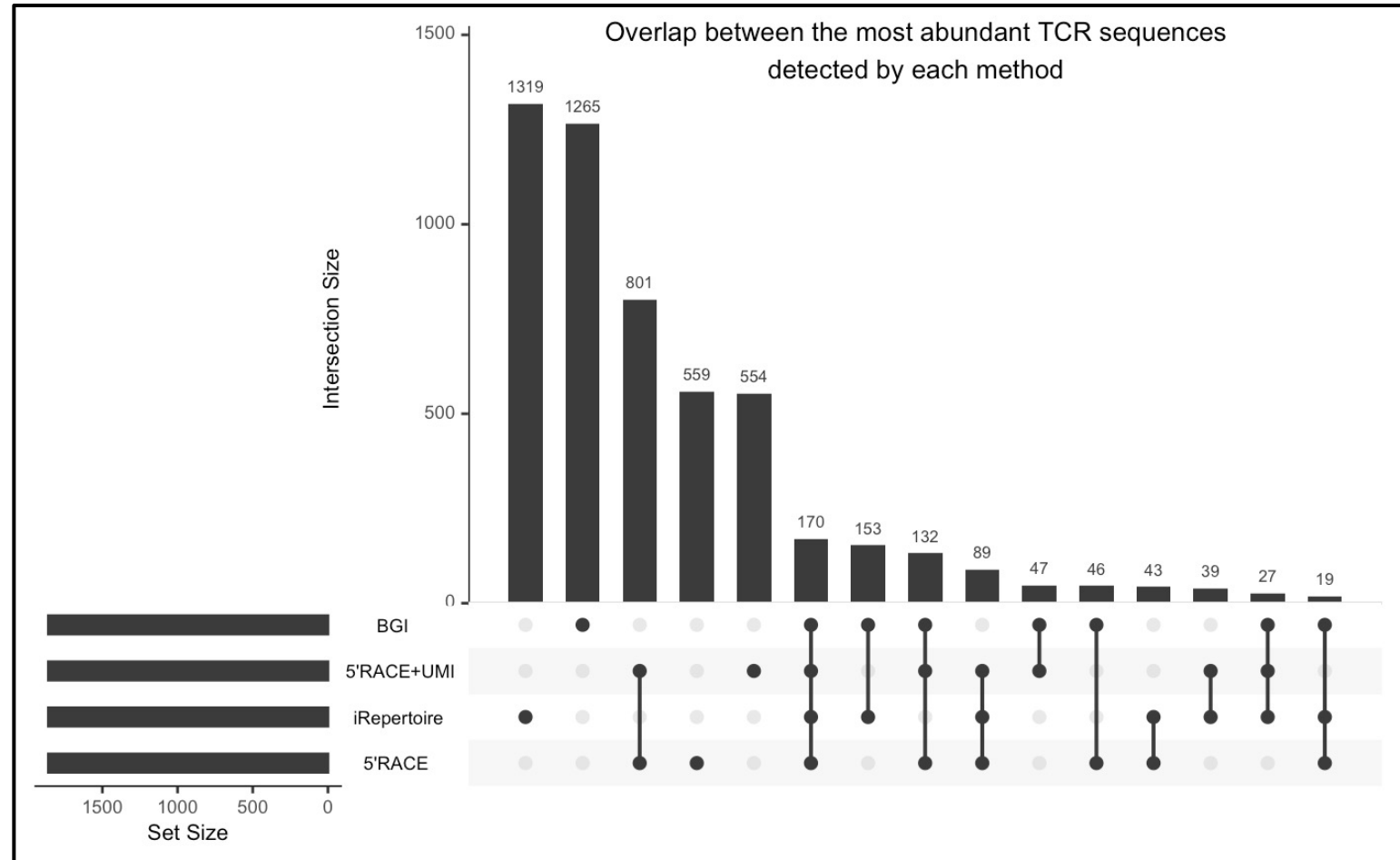# Visualizing Intersections of Multiple Sets: Comparing Venn diagrams and UpSet plots
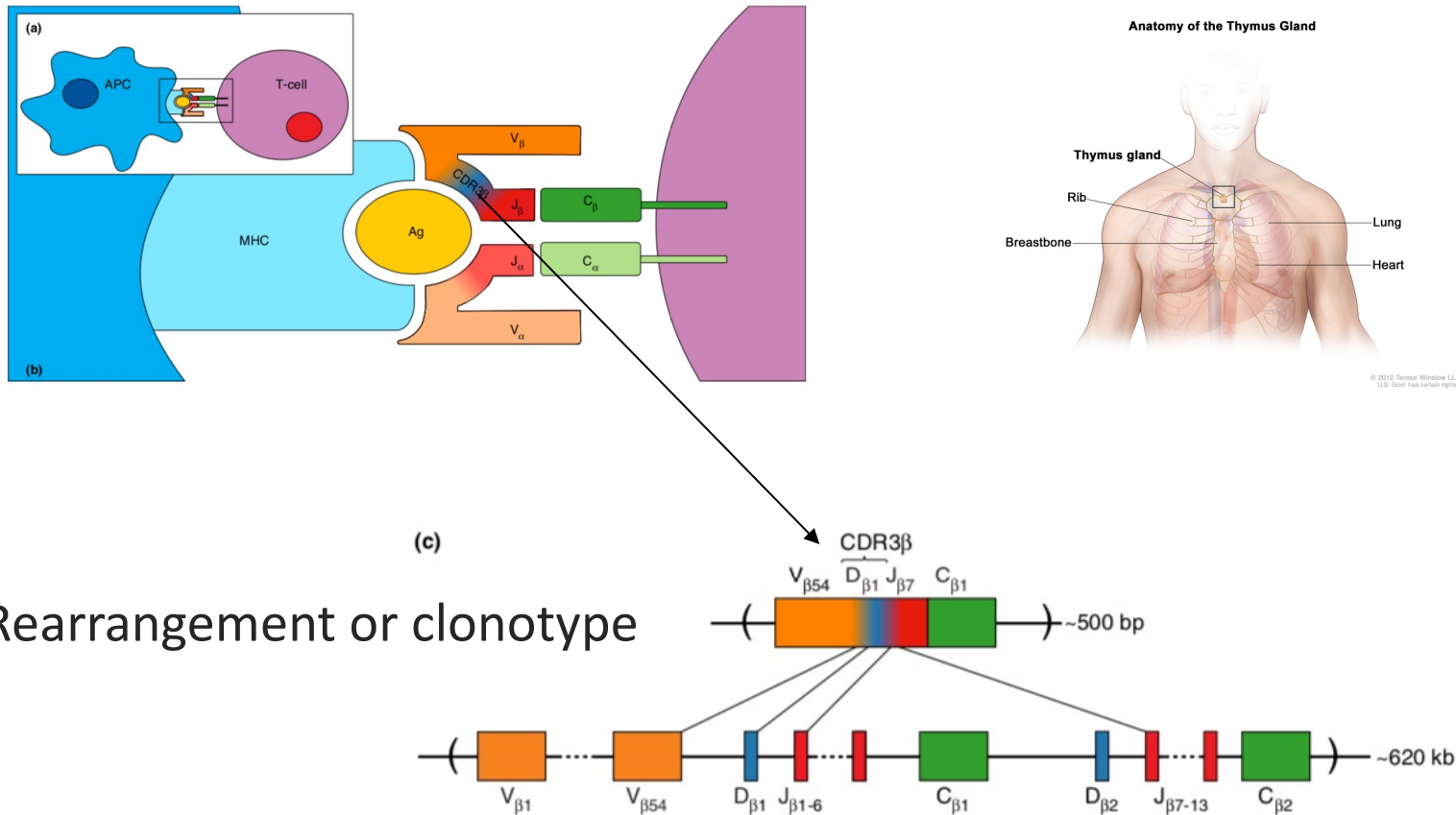
David Ross

BMI 525

9 June 2021

OHSU

The T-cell receptor (TCR) provides specificity to peptide-MHC complex

Rearrangement or clonotype

[Woodsworth et al. 2013, Terese Winslow NCI]

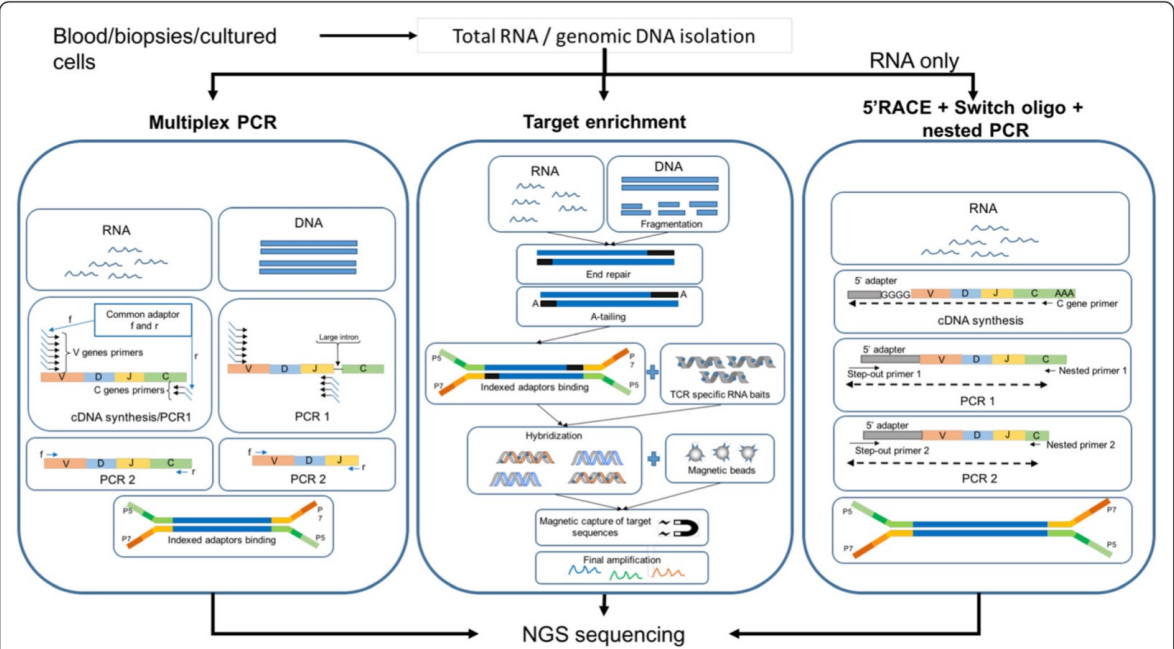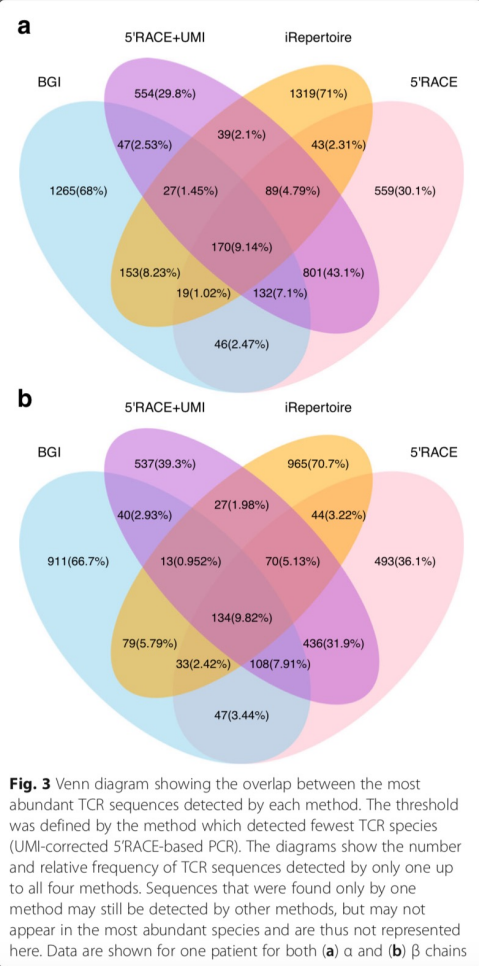# A methods paper evaluated different approaches for generating TCR repertoire data

Fig. 2 Exemplary workflow of three principal methodologies for TCR library preparation. The figure depicts a simplified workflow of the library preparation procedure using multiplex PCR, targeted in-solution enrichment and 5'RACE-switch-oligo nested PCR. Multiplex PCR is suitable for both RNA and gDNA sequencing. Samples undergo cDNA synthesis and 1 or more PCR steps followed by adaptor ligation and sequencing. While the forward primers for cDNA synthesis are designed to cover all known V genes for both starting materials, the location and number of the reverse primers differs, due to introns in DNA. Target enrichment, also applicable to both gDNA and RNA, is preceded by a standard library preparation including fragmentation for gDNA or mRNA purification for RNA, followed by end-repairing, A-tailing and finally adaptor ligation. The enrichment of target sequences is then performed using RNA baits complementary to the sequence of interest. The RNA baits hybridize with molecules in the library, which are then retrieved using magnetic beads and can undergo further amplification before sequencing. Nested PCR based on the 5'RACE and switch-oligo approach (only for RNA) makes use of the incorporation of an adaptor molecule at the 5' end of the cDNA during cDNA synthesis. The forward primer for a subsequent PCR is designed to bind to the 5' adaptor sequence, while the reverse primer is designed to bind to the C-region of the transcript. Hence, only one primer pair is required to cover the complete spectrum of possible V genes. Subsequent nested PCRs performed in the same fashion may increase outcome specificity. Finally, adaptor ligation is performed. The procedures showed in this picture constitute only an example of the different available methods

Fig. 3 Venn diagram showing the overlap between the most abundant TCR sequences detected by each method. The threshold was defined by the method which detected fewest TCR species (UMI-corrected 5'RACE-based PCR). The diagrams show the number and relative frequency of TCR sequences detected by only one up to all four methods. Sequences that were found only by one method may still be detected by other methods, but may not appear in the most abundant species and are thus not represented here. Data are shown for one patient for both (**a**) α and (**b**) β chains
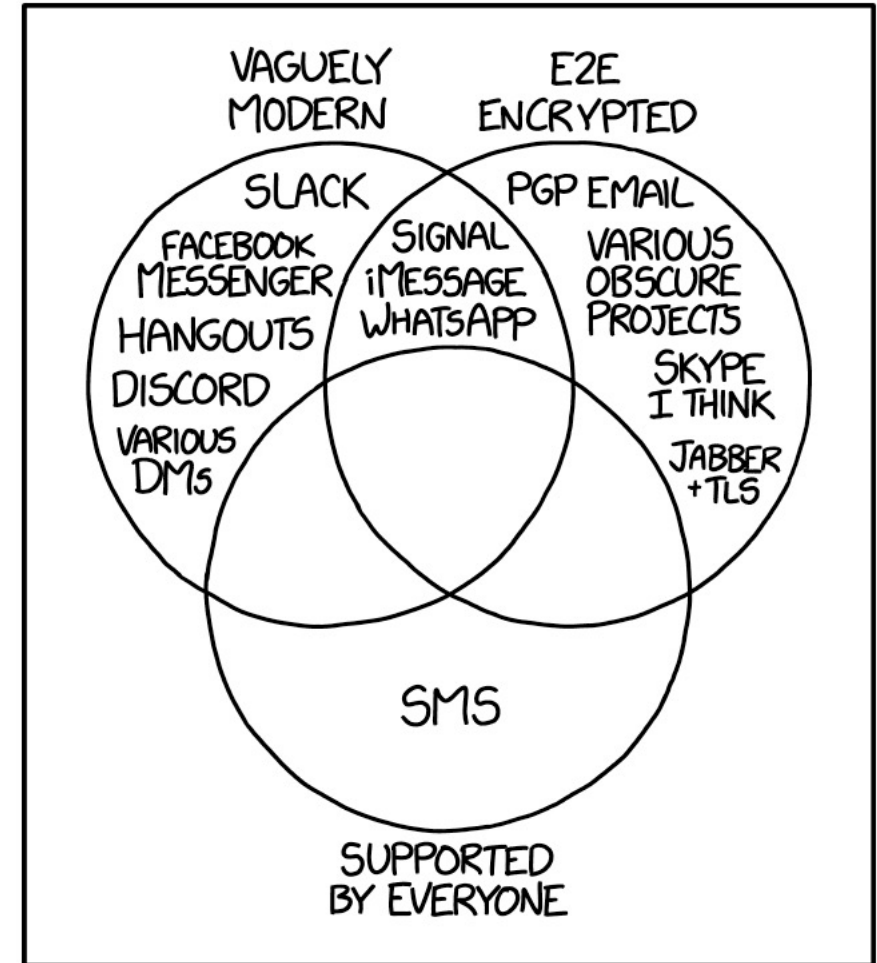
Rosati E, Dowds CM, Liaskou E, Henriksen EKK, Karlsen TH, Franke A. Overview of methodologies for T-cell receptor repertoire analysis. BMC Biotechnol. 2017 Jul 10;17(1):61.

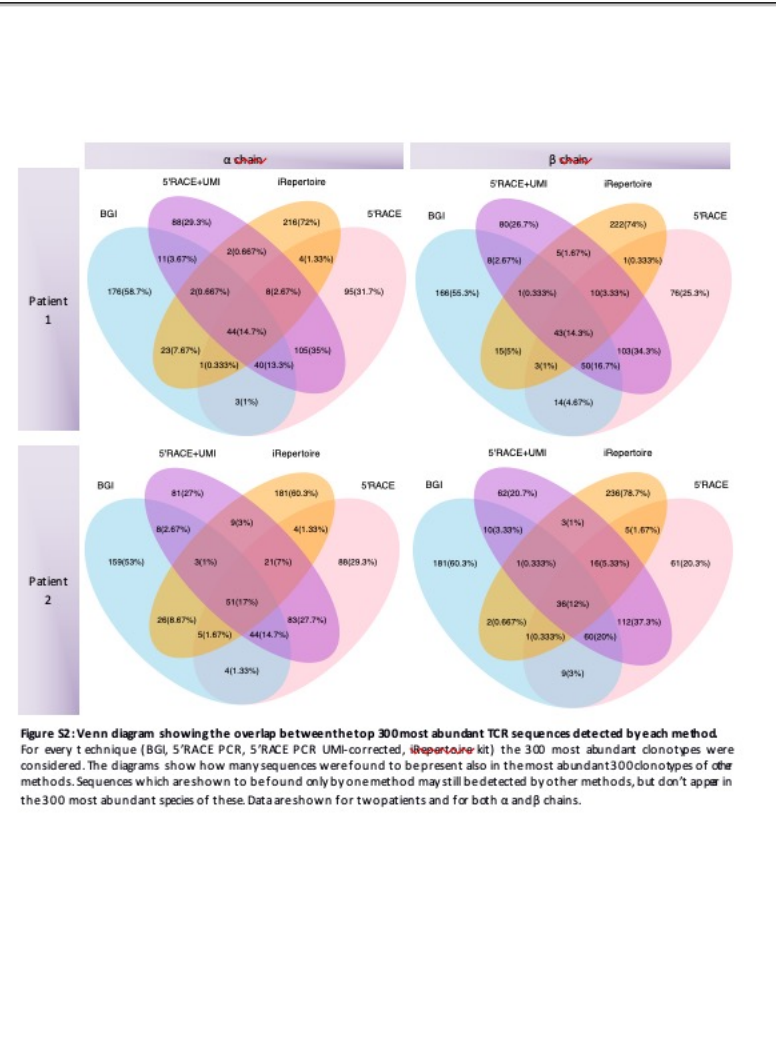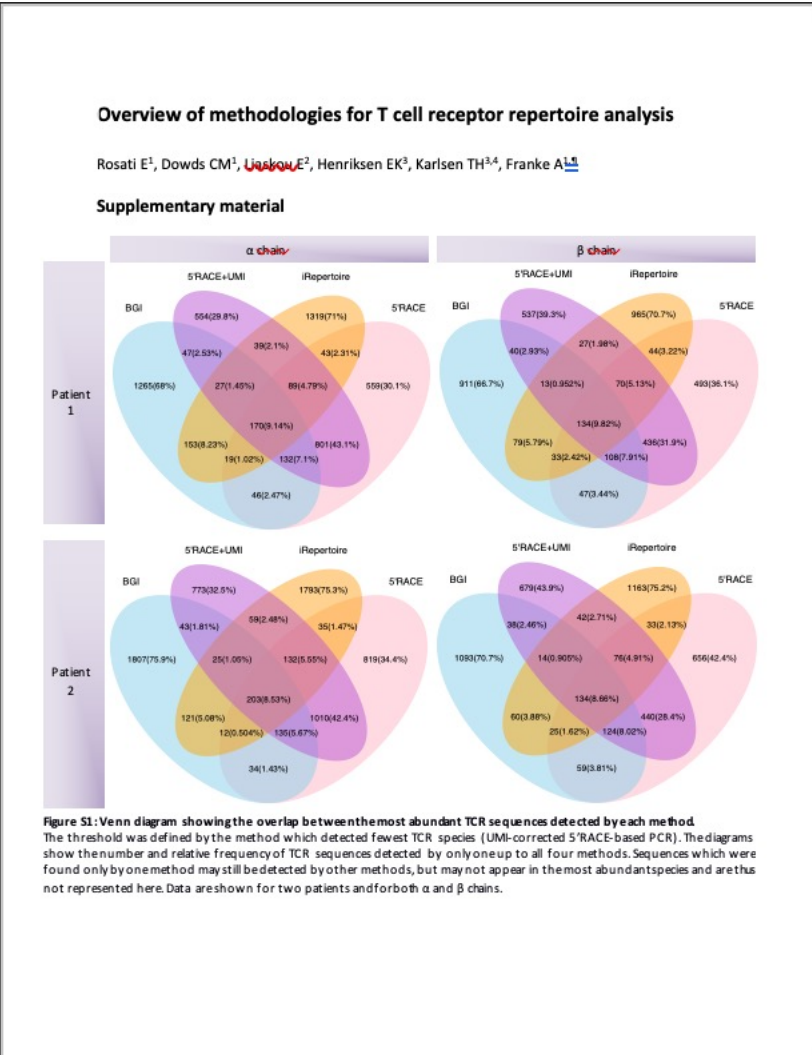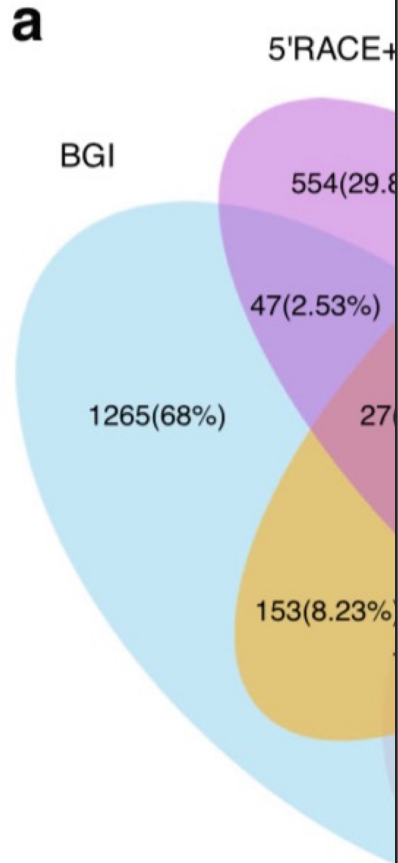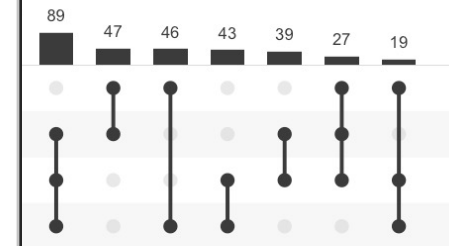# Venn diagrams are effective at showing relationships between two to three different sets

https://xkcd.com/773/
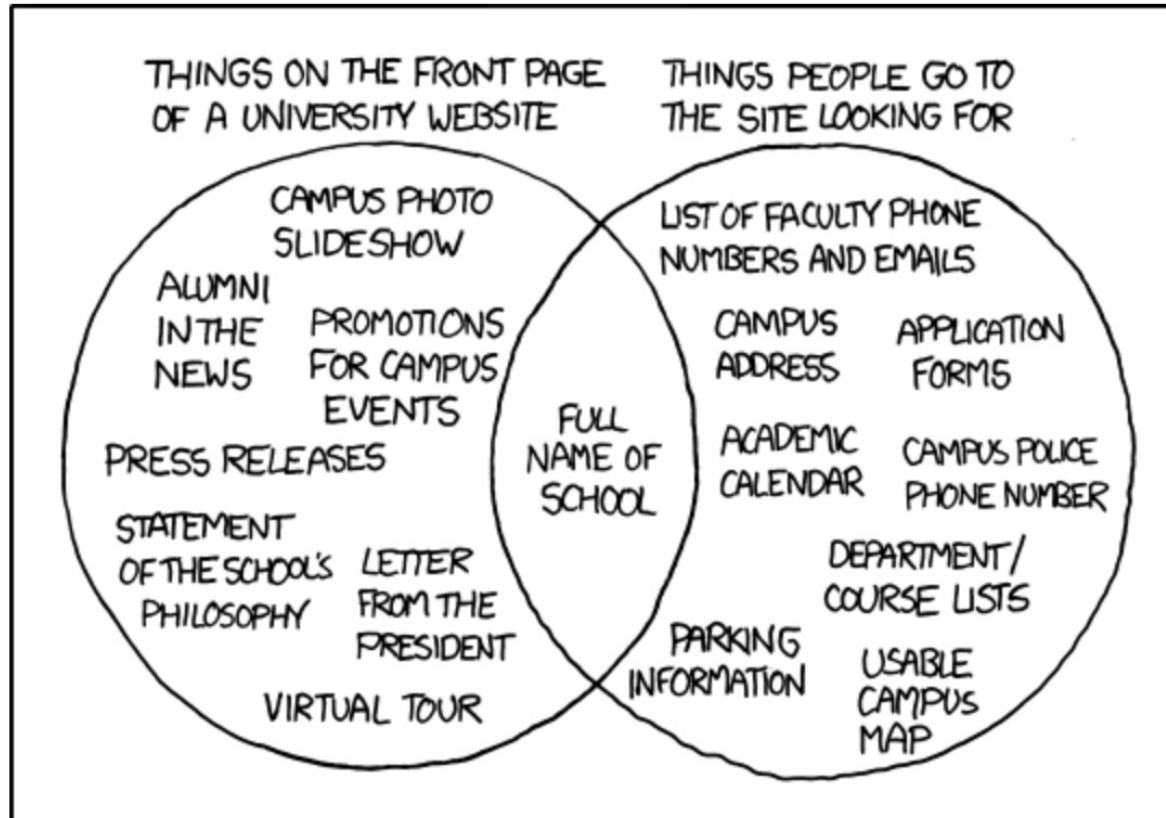
WHY SMS REFUSES TO DIE

https://xkcd.com/2365/

Analyzing relationships among greater than three sets is challenging with a Venn diagram.

UpSet plot is also available as an interactive tool

http://vcg.github.io/upset/