**Description and usage of cattle FAETH (v1.1) score**

The Functional-And-Evolutionary Trait Heritability (FAETH) score is a combined measure of sequence variant functionality (e.g., as expression eQTLs or under ChIP-seq peaks), evolutionary significance (e.g., as conserved sites or selection signatures) and of predicted heritability estimated from 34 complex traits of up to 50,000 cattle with over 17 million sequence variants.

Updates:

A gzip file of demo data ‘faeth.demo.tar.gz’ has been included in July 2019. This file can be used together with the online tutorial (https://ruidongxiang.com/2019/07/19/calculation-of-faeth-score/) to calculate FAETH score.

The FAETH score v1.1 is updated in June 2019. This update opens up additional 13 informative genome partitions that make of the FAETH score as described below.

Usage: The text file can be downloaded and unzipped. The overall FAETH score (column 4 or column 5, or the average of the two) can be used as a weight for each variant. You may use the overall FAEH score to rank variants newly discovered or weigh variants for making GRM. We recommend using calculated FAETH score or the average of the per-variant heritability across multiple genome partitions to rank variants. It is expected that the more functional/evolutionary data is included, the higher accurate the FAETH score will be. You may also use them as priors for genomic selection models such as BayesRC.

The updated FAETH score (column 6-20) also allows users to annotate their own variants with our FAETH data. For example, one can use the FAETH data to check whether a SNP he/she is interested is a splicing sQTL, under elocutionary constraint or related to coding activities.

Validation: The FAETH score is validated using independent datasets (Danish cattle).

Data specification: the downloadable FAETH score (text file with 17 million + rows x 20 columns) is a zipped (.gz) file which contains a header. The top lines in the file look like following:



From the 1st column to the 5th column, there are:

1. sequence variant ID

2. chromosome number

3. variant position

4. FAETH score based on bull data (bull.FAETH1.1) which are the average of the per-variant heritability of 13 genome partition for bulls (with bull.ldscore50kb and bull.maf)

5. FAETH score based on cow data (cow.FAETH1.1) which are the average of the per-variant heritability of 13 genome partition for cows (with cow.ldscore50kb and cow.maf).

The bull and cow FAETH score are very similar because except the MAF and LD score genome partitions, other partitions were identical between bulls and cows. You can choose to average them or use them separately. To avoid potential bias towards specific functional and/or evolutionary genome partitions, we recommend using the average of the per-variant heritability across multiple genome partitions to rank variants. However, you may also choose your own selection of the genome partitions to rank your variants.

From the 6th - 20th column, there are additional annotation for each variant, which are the per-variant heritability of:

6. annotation (7 categories/values, based on Variant Effect Prediction)

7. allele specific expression aseQTL (2 categories/values, i.e., aseQTL and the rest)

8. exon expression eeQTL (2 categories/values)

9. gene expression geQTL (2 categories/values)

10. bull.ldscore50kb (4 quartiles/values of LD score of 50 kb windows in bulls)

11. bull.maf (4 quartiles/values of minor allele frequency in bulls)

12. ChIPSeq (2 categories/values)

13. metabolites mQTL (2 categories/values)

14. selection.sig (selection signature between dairy and beef cattle, 2 categories/values)

15. young (variants that are recently selected, 2 categories/values)

16. variant.density50kb (4 quartiles/values of number of variant in 50kb windows)

17. splicing sQTL (2 categories/values)

18. conserved100way (variants under genomic sites conserved across 100 vertebrate species, 2 categories/values)

19. cow.ldscore50kb (4 quartiles/values of LD score of 50 kb windows in cows)

20. cow.maf (4 quartiles/values of minor allele frequency in cows)

For details of these genome partitions, please refer to Table 1, Results and Methods in the paper (reference below). The following key will help to find the per-variant heritability values for different categories within each one of the 13 partitions (as shown in Figure 3 in the paper).

6. annotation: noncoding.related > Splice.sites > UTR > geneend > merged.coding > intergenic\_variant > intron

7. aseQTL: aseQTL > rest

8. eeQTL: eeQTL > rest

9. geQTL: geQTL > rest

10. bull.ldscore50kb: quartile 4 > quartile 3 > quartile 2 > quartile 1

11. bull.maf: quartile 4 > quartile 3 > quartile 2 > quartile 1

12. ChIPSeq: ChIPSeq > rest

13. mQTL: mQTL > rest

14. selection.sig: selection.sig > rest

15. young: young > rest

16. variant.density50kb: quartile 4 > quartile 3 > quartile 1 > quartile 2

17. sQTL: sQTL > rest

18. conserved100way: conserved100way > rest

19. cow.ldscore50kb: quartile 4 > quartile 3 > quartile 2 > quartile 1

20. cow.maf: quartile 4 > quartile 3 > quartile 2 > quartile 1

The meaning of these keys for SNP annotation:

For example, at the 6th column, you will find 7 numeric unique values, representing 7 types of variants in this category. These values from big to small represent noncoding.related, Splice.sites, UTR, geneend, merged.coding, intergenic\_variant and intron. Therefore, SNPs corresponding to these values can be annotated accordingly; e.g., SNPs with the largest values in column 6 are noncoding.related, and SNPs with the 2nd largest values in column 6 are Splice.sites and etc. At the 7th column (aseQTL), only 2 unique values are there. This means that those SNPs with the larger values at the 7th column are identified as aseQTLs and the other SNPs are not aseQTLs because they have the smaller values at the 7th column.

Citation:

Xiang, R., Van Den Berg, I., MacLeod, I. M., Hayes, B. J., Prowse-Wilkins, C. P., Wang, M., ... Goddard, M.E. (2019). Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *PNAS (2019), In Press*.