

# Supplementary materials

## Signal metrics analysis of oscillatory patterns in bacterial multi-omic networks

Francesco Bardozzo, Pietro Lió and Roberto Tagliaferri

November 2, 2020

### Abstract

This document contains all the supplementary material related to the paper: *Signal metrics analysis of oscillatory patterns in bacterial multi-omic networks* - Bardozzo F., Lió P. and Tagliaferri R. The sections in this document report tables, graphs, as well as correctness theorems of the proposed algorithms. In detail, all the parts of code in R language to obtain multi-omic oscillations on signals are linked in the respective sections. In addition on the GitHub repository <https://github.com/lodeguns/Multi-omicSignals> there are two working examples to test our change point detection algorithms and further, for finding oscillations on your own biological sequences.

Should you need help running our code, please contact: [fbardozzo@unisa.it](mailto:fbardozzo@unisa.it)

### Content Index:

- **Section 1:**  
Inter-organisms amplitude consensus procedure - IOAC procedure.
- **Section 2:**  
Statistics between organisms and for every multi-omic combination on the half-periodic features ( $\theta$ s)
- **Section 3:**  
Theorem and Proof of correctness of the change point detection algorithms
- **Section 4:**  
Computation of the MAE for the  $osc_s$  samples perturbed by additive noise and random shuffled
- **Section 5:** Implementation of the oscillation metrics  $osc_k$  and  $osc_s$
- **Section 6:** Whole dataset of the multi-omic signal metrics
- **Section 7:**  
Network-level synchronizations of the multi-omic oscillations
- **Section 8:** Tables of  $\theta$ s

## 1 Inter-organisms amplitude consensus (IOAC) procedure

An amplitude discretization process is applied in order to compare the signals between different organisms. The number of levels of discretization  $N$  is decided through the following procedure, that it is called: [Inter Organisms Amplitude Consensus \(IOAC\)](#). Several experiments were carried out on the type of statistical distribution of the signal values and on the different methodologies applicable ( according to the distribution) to identify the correct number of bins to use for the discretization. The research led to the use of Doane’s formula as explained in the paper. In particular, the number of most frequent bins for all omic values arranged on the signals extracted from all organisms is 9. Other values have also been identified, from 7 at 11 but with much lower relative frequencies as shown in Table 1.

Table 1. The IOAC procedure, comparing all the single omics, estimates  $N$  equal to 9 (bold column)

	$N = 7$	8	<b>9</b>	10
CC	1.32%	1.42%	<b>76.38%</b>	20.85%
CAI	8.33%	8.33%	<b>41.67%</b>	41.67%
MW	8.33%	8, 33%	<b>58.33%</b>	25.00%

Figure 1: Table 1

## 2 Median change point detector in order to search the half-periodic windows lenght( Algorithm 1) - Step 2

Before developing metrics that studied the oscillation we made sure to verify on the entire dataset that the periodic characteristics (on their median values) and that on the boundary conditions (min and max periods) really existed. The median change point detection algorithm for the estimation of the half periodicity search windows ( $\theta$ ) is calculated as described [here](#). See also the table of the estimated  $\theta$ s applied for all the organisms and for all the multi-omic combinations in section 8 or [here](#).

### 3 Theorem 1

The Algorithm 2 gives in output an  $osc_s$  index defined as:

$$osc_s = \frac{\sum_{i=1}^d m\vec{v}q[i] \cdot m\vec{v}l[i]}{(N-1) \sum_{i=1}^d m\vec{v}l[i]} \quad (1)$$

equal to 1 if and only if ( $\Leftrightarrow$ ) the observed signal presents a perfect oscillation.

Proof. of correctness:

1. Proof of the ( $\Rightarrow$ )

Given  $\mathbf{S}$  as the set of the multi-omic signals, as they are defined in the paper. Given a set of classes:  $\mathbf{N} := \{0, 1, 2, \dots, N-1\}$ , as they are specified in the paper, in which the observed signal is discretized. Assuming that a signal  $\vec{x} \in S$  has a perfect oscillation, means that  $\vec{x}$  is defined as:

$$\vec{x} = [\min(N), \max(N), \min(N), \dots, \max(N), \min(N), \max(N)] = [0, N-1, 0, \dots, N-1, 0, N-1] \in \mathbf{S} \quad (2)$$

of wavelength equal to  $d = |\vec{x}| - 1$ .

Since  $\vec{x}$  has a perfect oscillation, it presents  $m\vec{v}l$  of all ones:

$$m\vec{v}l = [1, 1, 1, 1, 1, \dots, 1]_d \quad (3)$$

and, obviously, it presents a  $m\vec{v}q$  of all values equal to  $N-1$ :

$$m\vec{v}q = [N-1, N-1, \dots, N-1]_d \quad (4)$$

By their definition, the dot product between  $m\vec{v}l$  and  $m\vec{v}q$  is equal to:

$$m\vec{v}q \cdot m\vec{v}l = \sum_{i=0}^d m\vec{v}q[i] m\vec{v}l[i] = d(N-1) \quad (5)$$

Replacing 5 in 1 we obtain:

$$osc_s = \frac{d(N-1)}{(N-1) \sum_{i=1}^d m\vec{v}l[i]} = \frac{d}{\sum_{i=1}^d m\vec{v}l[i]} \quad (6)$$

Evaluating the denominator of 6 we obtain:

$$\sum_{i=1}^d m\vec{v}l[i] = d \quad (7)$$

So the  $osc_s$  index for  $\vec{x}$  is necessary equal to 1.  $\square$

2. Proof of the ( $\Leftarrow$ ):

Now we will assume that a signal  $\vec{x}$  has an  $osc_s$  index equal to 1 and we will prove that  $\vec{x}$  has a perfect oscillation. By this assumption, we state that:

$$osc_s = \frac{\sum_{i=1}^d m\vec{v}q[i] \cdot m\vec{v}l[i]}{(N-1) \sum_{i=1}^d m\vec{v}l[i]} = 1 \quad (8)$$

In order to have a fraction equal to 1, the denominator and the numerator must be equal:

$$\sum_{i=1}^d m\vec{v}q[i] \cdot m\vec{v}l[i] = (N-1) \sum_{i=1}^d m\vec{v}l[i] \quad (9)$$

Developing the left side of 9 we obtain:

$$\sum_{i=1}^d m\vec{v}q[i] \cdot m\vec{v}l[i] = m\vec{v}q[0] m\vec{v}l[0] + m\vec{v}q[1] m\vec{v}l[1] + \dots + m\vec{v}q[d] m\vec{v}l[d] \quad (10)$$

The only way to respect the equality defined in 9 is to assume that  $m\vec{v}q$  is a constant vector, precisely equal to  $N-1$ . This means that, if a signal has a vector  $m\vec{v}q$  constant and equal to  $N-1$ , so the signal has a perfect oscillation, by the definition of  $m\vec{v}q$ .  $\square$

## 4 Computation of the MAE for the $osc_s$ samples perturbed by noise and random shuffled

We define 3 samples of  $osc_s$  index of length  $n$ . The first one  $\Omega_1$  is a vector of  $osc_s$  computed on the original signals. The second one  $\Omega_2$  is a vector of  $osc_s$  computed on the noisy signals. The third one  $\Omega_3$  is a vector of  $osc_s$  computed on the random shuffled elements of the original signals.

The dissimilarities are computed between  $\Omega_1$  vs  $\Omega_2$  and between  $\Omega_1$  vs  $\Omega_3$ .

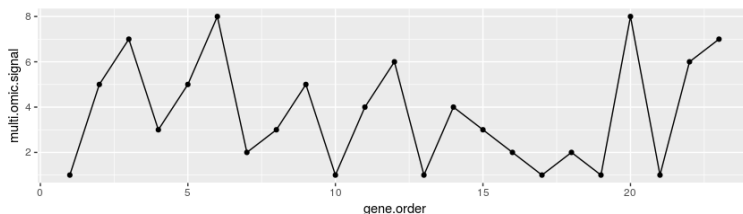
The  $\Omega$  vectors difference measurements may be scaled to a unit vector by:

$$MAE(\Omega_1, \Omega_{2|3}) = \frac{1}{n} \|\Omega_1 - \Omega_{2|3}\|_1 = \frac{1}{n} \sum |\Omega_{1_i} - \Omega_{2|3_i}|$$

which is known as Mean-Absolute Error (MAE). The values of MAE between these three subsets are reported in the paper.

## 5 Implementation of the oscillation metrics.

The  $osc_s$  index and  $osc_k$  index for multi-omic spatial signal were calculated with the script linked [here](#). See also the next subsection for a whole dataset of these estimated values.



## 6 Whole dataset of the multi-omic signal metrics

Note, in order to make the algorithms described above faster and easier to use, we have provided in the source code some simple examples. The same algorithms, however, have been applied to more than two million of multi-omic signals and the results obtained are collected in this dataset: [global.nt](#). In addition, the dataset already contains all the associations to the functional classes of [KEGG Orthology](#) and to the identifiers linked to the [COLOMBOS v3.0](#) condition contrasts. More in detail, below we describe the fields of the dataset:

- *score* : the  $osc_s$  index,
- *m.s* : the average value of  $osc_s$

- *med.s* : the median value of  $osc_s$
- *sd.s* : the standard deviation of  $osc_s$
- *m.w* : the average value of the period lengths
- *med.w* : the median value of the period lengths
- *sd.w* : the standard deviation of the period lengths
- *change.w* : number of periodic oscillations
- *v.change.w* :  $osc_k$  index
- *path.l* : length of the signal
- *n.path* : KEGG pathway ID
- *exp.cr* : COLOMBOS v3.0 condition contrast ID
- *exp.ref* : COLOMBOS v3.0 treatment experiment ID
- *exp.ctr* : COLOMBOS v3.0 control experiment ID
- *kegg.id* : KEGG organism ID
- *code* : ID of the multi-omic combination
- *class* : KEGG orthology level 1
- *func* : KEGG orthology level 2
- *pathwaymap* : KEGG orthology level 3

Note, for the field *code*, these are the possible IDs:

- *n1* : CAI + Molecular Weight
- *n2* : CAI + mRNA CCs
- *n3* : Molecular Weight + mRNA CCs
- *n4* : CAI + Molecular Weight + mRNA CCs

While, for the operon compressed signals:

- *o1* : CAI + Molecular Weight
- *o2* : CAI + mRNA CCs
- *o3* : Molecular Weight + mRNA CCs
- *o4* : CAI + Molecular Weight + mRNA CCs

## 7 Network-level synchronizations and plots

Once the dataset described above has been downloaded, it will be possible to analyze the [network-level synchronizations](#). To do this, in this section we provide [script1](#) and [script2](#). Through these scripts it is possible to interact with the signal dataset and perform your own analyses in order to identify the oscillatory networks, for example by considering either a restricted subset of organisms either only specific COLOMBOS v3.0 experiments. However, in order to speed up the visualization of the plotted figures, all the intersections of the oscillatory networks with  $\phi = 0.8$  have been generated into pdf files. The analysis was done [between bacteria](#) and [within bacteria](#). Furthermore, **the images related to between bacteria synchronizations are provided at the end of this document**. In the respective section of GitHub you can find on the readme the pdf filenames structure in order to identify the organisms, the multi-omic combinations and KEGG orthology levels. The relative dependencies to the RData datasets are explicitly indicated in the source code.

## 8 Tables of $\theta$ s estimation

Final between-organisms table of  $\theta$ s from the estimated  $\hat{W}L$  with Algorithm 1

MOC	$\theta$	Signal
CAI + MW	5	standard
CAI + MW	5	operon compressed
CAI + CCs	6	standard
CAI + CCs	6	operon compressed
CAI + MW	9	standard
CCs + MW	7	operon compressed
CCs + MW + CAI	7	standard
CCs + MW + CAI	6	operon compressed

$\hat{W}L$  : CAI + Molecular Weight (without operon compression)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3	3.026883	0.7526783	1	5
bsu	3	3.120692	0.7804005	1	5
bth	3	3.292180	0.8541587	1	5
cac	3	3.086670	0.8015822	1	5
cje	3	2.897637	0.7430687	1	4
eco	3	3.041668	0.7876022	1	5
hpy	3	3.016070	0.9288832	1	5
mtu	3	3.132187	0.9080303	1	5
pae	3	3.103962	0.6305740	1	4
sme	3	3.159035	0.8513161	1	5
stm	3	3.098904	0.6962391	1	4
<b>MEANS</b>	3.0000000	3.0887172	0.7940485	1.0000000	4.7272727



$\hat{W}L$  : CAI + CCs Expression (without operon compression)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3	3.077475	0.7906031	1	6.0
bsu	3	3.123336	0.7834352	1	6.0
bth	3	3.238475	0.8271141	1	5.5
cac	3	3.055546	0.7866987	1	5.0
cje	3	2.854868	0.7150365	1	5.0
eco	3	3.009572	0.7454507	1	5.0
hpy	3	2.939054	0.8684818	1	5.5
mtu	3	3.124221	0.8481821	1	5.5
pae	3	3.087685	0.6300326	1	6.0
sme	3	3.155368	0.8690302	1	6.0
stm	3	2.997901	0.6387462	1	6.0
<b>MEANS</b>	3.0000000	3.0603183	0.7729828	1.0000000	5.5909091

 $\hat{W}L$  : Molecular weight + CCs Expression (without operon compression)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3	3.289308	0.9083618	1	9.0
bsu	3	3.126198	0.8125143	1	8.0
bth	4	3.455948	0.9411688	1	6.0
cac	3	3.242042	0.9090368	1	7.0
cje	3	2.957530	0.8127884	1	5.0
eco	3.5	3.343877	0.9513413	1	6.5
hpy	3	3.142207	1.0	1	6.0
mtu	3	3.272489	0.9441163	1	6.5
pae	3	3.369740	0.7990332	1	6.0
sme	3	3.168847	0.8679407	1	6.0
stm	3.5	3.396924	0.8822160	1	6.0
<b>MEANS</b>	3.1818182	3.2513736	0.8966079	1.0000000	6.5454545

$\hat{W}L$  : CAI + Molecular Weight + CCs Expression ( without operon compression)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3	3.051562	0.7660589	1	6.0
bsu	3	3.114014	0.7809811	1	6.5
bth	3	3.269988	0.8346170	1	5.0
cac	3	3.072169	0.7948275	1	6.0
cje	3	2.892553	0.7453163	1	4.5
eco	3	3.044299	0.7812696	1	5.0
hpy	3	3.010309	0.9176502	1	6.0
mtu	3	3.118348	0.8598279	1	5.5
pae	3	3.125488	0.6515653	1	6.0
sme	3	3.162967	0.8648532	1	6.0
stm	3	3.093292	0.6991620	1	5.0
<b>MEANS</b>	3	3.0868171	0.7905572	1.0000000	5.5909091

#### Window estimation on operon compression

$\hat{W}L$  : CAI + Molecular Weight ( Not jet operon compressed signals and after compression)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3	3.142872	0.5288159	1.0	4
bsu	3	3.115355	0.6401679	1.0	4
bth	3.5	3.538462	0.6032805	3.0	5
cac	3	3.135145	0.7037910	1.0	5
cje	3	2.666667	0.5376018	1.0	3
eco	3	3.197368	0.5323805	2.0	5
hpy	3.5	3.411765	0.6000174	2.0	4
mtu	3	3.173052	0.3654015	3.0	4
pae	3	3.138897	0.4185364	2.0	4
sme	3	3.187500	0.7190200	1.0	5
stm	3	3.300020	0.5237524	2.5	5
<b>MEANS</b>	3.0909091	3.1824638	0.5611605	1.7727273	4.3636364

(compressed)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3.0	3.214307	0.6128470	1.0	5.0
bsu	3.0	3.365413	0.4287421	3.0	4.0
bth	4.0	3.673077	0.5360833	3.0	5.0
cac	3.0	3.351377	0.5310266	2.5	5.0
cje	3.0	3.233333	0.4028566	2.5	4.0
eco	3.0	3.421053	0.5795604	3.0	5.0
hpy	3.5	3.411765	0.6000174	2.0	4
mtu	3.5	3.529412	0.5278920	3.0	4.5
pae	3.0	3.230787	0.4213526	3.0	4.0
sme	3.0	3.138897	0.3458482	3.0	4.0
stm	3.0	3.375000	0.6081418	1.0	5.0
<b>MEANS</b>	3.1363636	3.3575165	0.4958348	2.5454545	4.5000000

(not operon compressed)

$\hat{W}L$ : CAI + CCs expression ( Not jet operon compressed and after compression)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3	3.268225	0.6162833	1.0	5.0
bsu	3	3.282599	0.7278886	1.0	5.0
bth	3	3.466505	0.5422313	3.0	6.0
cac	3	3.109837	0.6544316	1.0	5.0
cje	3	2.676974	0.5515709	1.0	5.0
eco	3	3.197368	0.5323805	2.0	5.0
hpy	3	3.188412	0.5437645	2.0	5.0
mtu	3	3.179168	0.3660061	3.0	5.0
pae	3	3.145425	0.4382088	2.0	5.5
sme	3	3.191739	0.7187863	1.0	6.0
stm	3	3.146978	0.3680019	2.5	5.0
<b>MEANS</b>	3	3.1684753	0.5508685	1.7727273	5.2272727

(compressed)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3.0	3.333401	0.6322089	1.0	5
bsu	3.5	3.502693	0.5561906	3.0	5
bth	3.5	3.528644	0.5167179	3.0	5
cac	3.0	3.332234	0.5030150	2.5	5
cje	3.0	3.099123	0.3365273	2.5	4
eco	3.0	3.308684	0.4651702	3.0	5
hpy	3.0	3.337461	0.4670644	3.0	5
mtu	3.0	3.318678	0.4631874	3.0	5
pae	3.0	3.204817	0.3964780	3.0	5
sme	3.0	3.388505	0.6554176	1.0	6
stm	3.0	3.162241	0.3592644	3.0	5
<b>MEANS</b>	3.0909091	3.3196801	0.4864765	2.5454545	5.0000000

(not operon compressed)

$\hat{W}L$  : Molecular weight and CCs expression( Not jet operon compressed and after compression)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3.5	3.454513	0.7097622	1.0	6.0
bsu	3.0	3.277983	0.8021867	1.0	7.0
bth	3.5	3.588762	0.6091726	3.0	6.0
cac	3.0	3.381560	0.8135864	1.0	6.0
cje	3.0	2.771272	0.6671862	1.0	5.0
eco	3.5	3.571868	0.6874662	2.0	6.0
hpy	3.5	3.483636	0.7436738	2.0	6.0
mtu	3.5	3.563813	0.5731475	3.0	6.0
pae	3.5	3.526335	0.6358427	2.0	6.0
sme	3.0	3.222202	0.7216577	1.0	6.0
stm	3.5	3.563852	0.6332037	2.5	6.5
<b>MEANS</b>	3.318182	3.400527	0.690626	1.772727	6.045455

(compressed)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	4	3.587187	0.7355230	1.0	9.0
bsu	3	3.515732	0.6455687	3.0	8.0
bth	4	3.858050	0.5997616	3.0	6.0
cac	4	3.648408	0.6222261	2.5	7.0
cje	3	3.357895	0.5371082	2.5	5.0
eco	4	3.820026	0.6438676	3.0	6.5
hpy	4	3.768465	0.6451115	3.0	6.0
mtu	4	3.774406	0.5763519	3.0	6.5
pae	4	3.694789	0.6218183	3.0	6.0
sme	3	3.442890	0.6604995	1.0	6.0
stm	4	3.767707	0.6315270	3.0	6.0
<b>MEANS</b>	3.727273	3.657778	0.629033	2.545455	6.545455

(not operon compressed)

$\hat{W}L$ : CAI + Molecular Weight + CCs Expression (not jet operon compressed and compressed)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3.0	3.206432	0.5920474	1.0	5.0
bsu	3.0	3.165667	0.6783979	1.0	5.0
bth	3.5	3.556422	0.6045234	3.0	5.0
cac	3.0	3.118010	0.6745034	1.0	5.0
cje	3.0	2.686184	0.5615729	1.0	4.0
eco	3.0	3.249579	0.5605360	2.0	5.5
hpy	3.0	3.332596	0.6141357	2.0	5.5
mtu	3.0	3.222940	0.4042796	3.0	5.0
pae	3.0	3.135424	0.4170666	2.0	5.5
sme	3.0	3.204215	0.7221253	1.0	6.0
stm	3.0	3.281218	0.4780867	2.5	5.0
<b>MEANS</b>	3.0454545	3.1962443	0.5733886	1.7727273	5.1363636

(compressed)

KEGG.ID	MEDIAN	MEAN	SD	MIN	MAX
bce	3.0	3.266710	0.6096960	1.0	6.0
bsu	3.0	3.370759	0.4706230	3.0	5.0
bth	4.0	3.662047	0.5309143	3.0	5.0
cac	3.0	3.361270	0.5218437	2.5	6.0
cje	3.0	3.217105	0.4144880	2.5	4.0
eco	3.0	3.395000	0.5457647	3.0	5.0
hpy	3.5	3.490712	0.5149555	3.0	5.5
mtu	3.0	3.250250	0.4191831	3.0	5.0
pae	3.0	3.179871	0.3790489	3.0	5.0
sme	3.0	3.395470	0.6451882	1.0	6.0
stm	3.0	3.361749	0.4813288	3.0	5.0
<b>MEANS</b>	3.1363636	3.3591768	0.5030031	2.5454545	5.2272727

(not operon compressed)

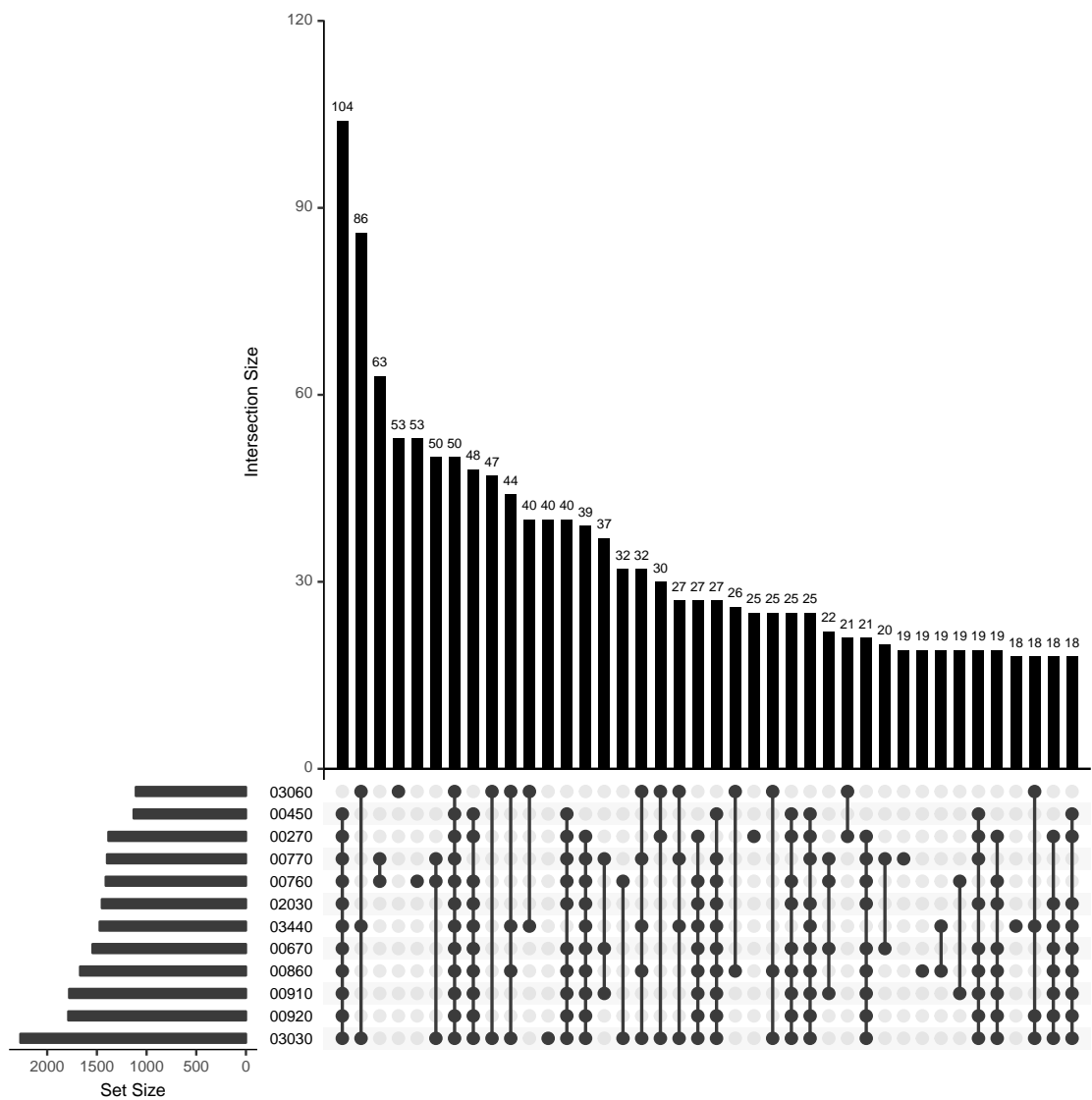


Figure 2: KO LEVEL: KEGG orthology level 1 (pathways) MOC: CAI + mRNA CCs

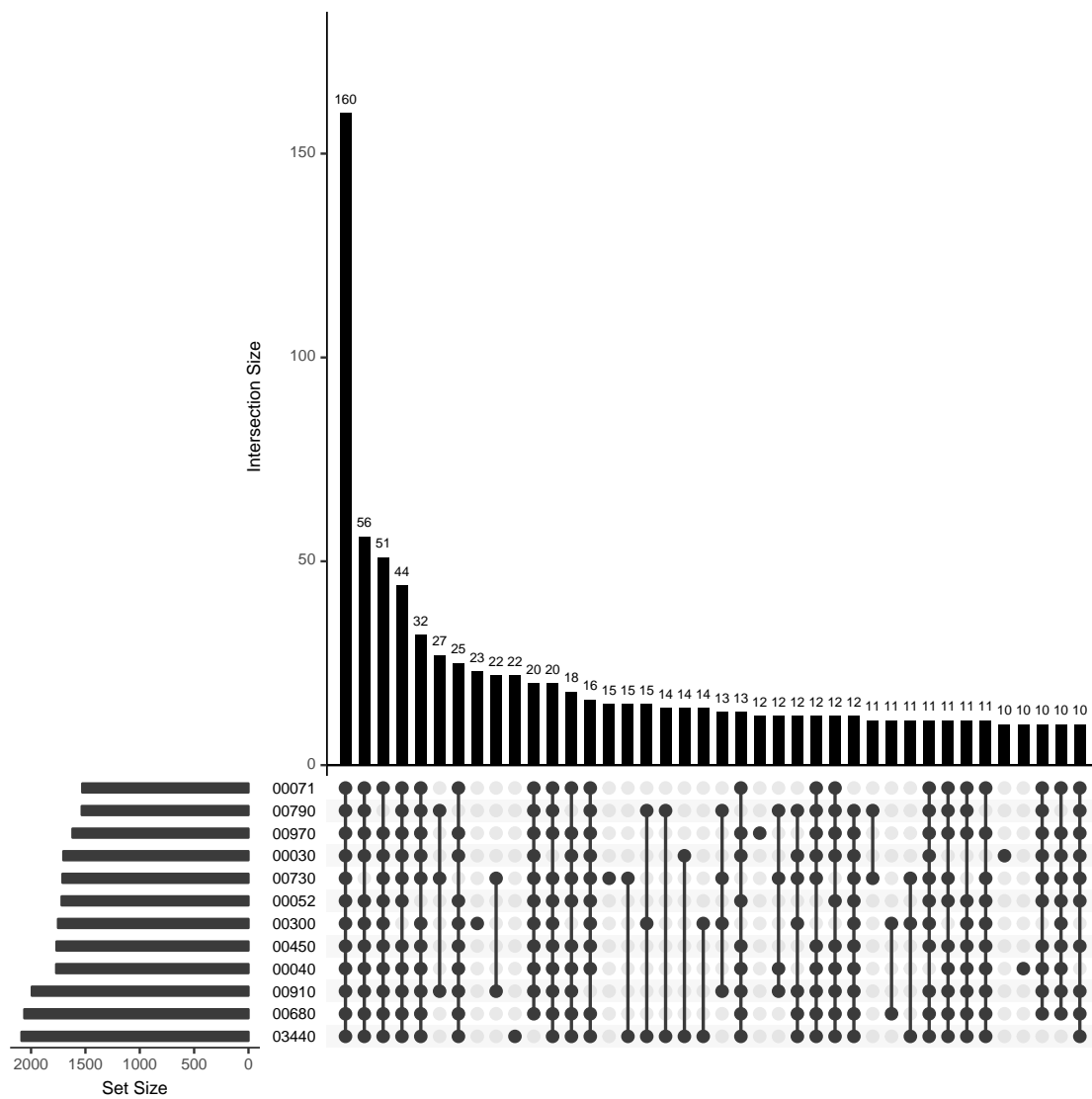


Figure 3: KO LEVEL: KEGG orthology level 1 (pathways) MOC: Molecular Weight + mRNA CCs

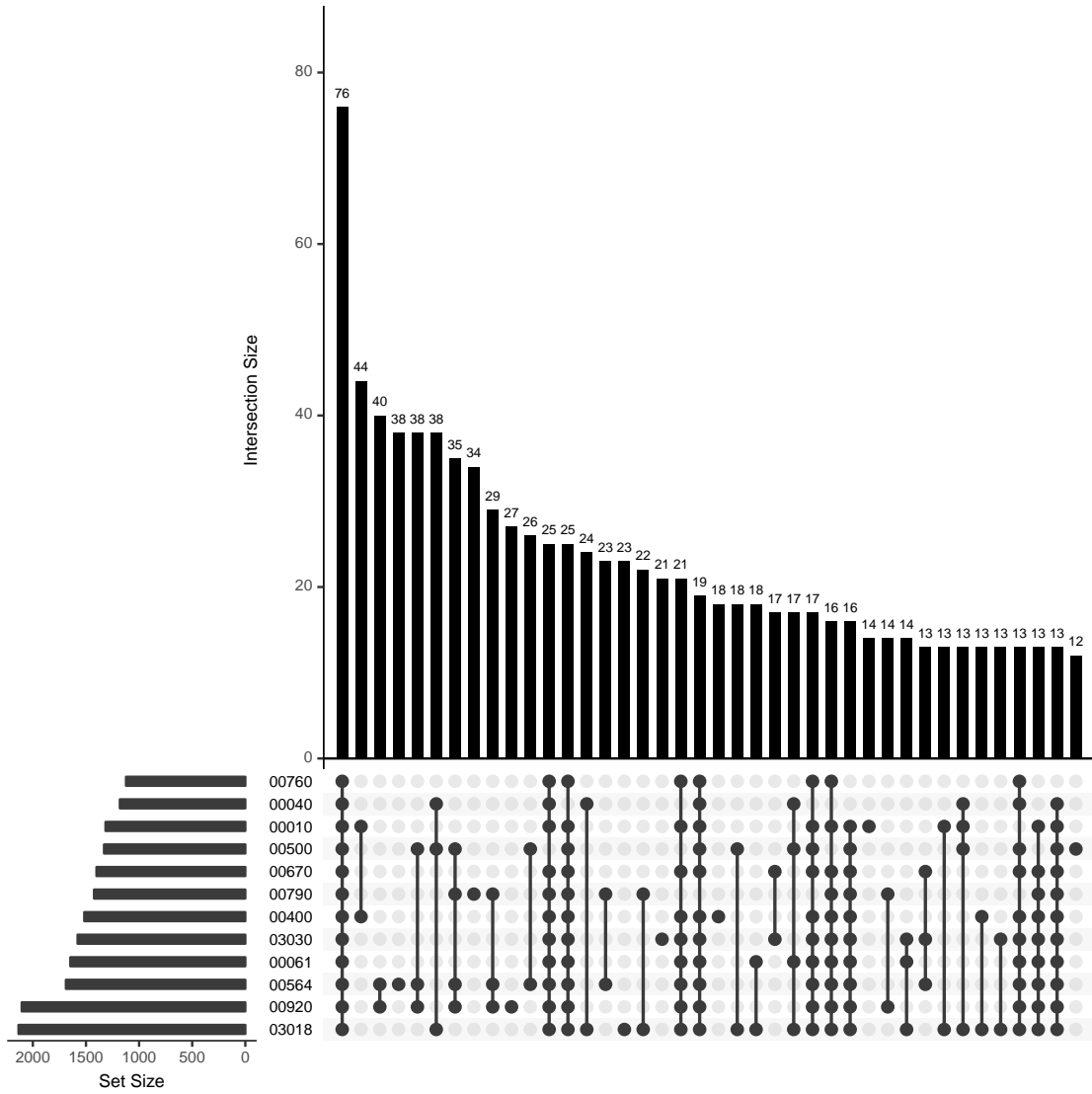


Figure 4: KO LEVEL: KEGG orthology level 1 (pathways) MOC: CAI + Molecular Weight + mRNA CCs



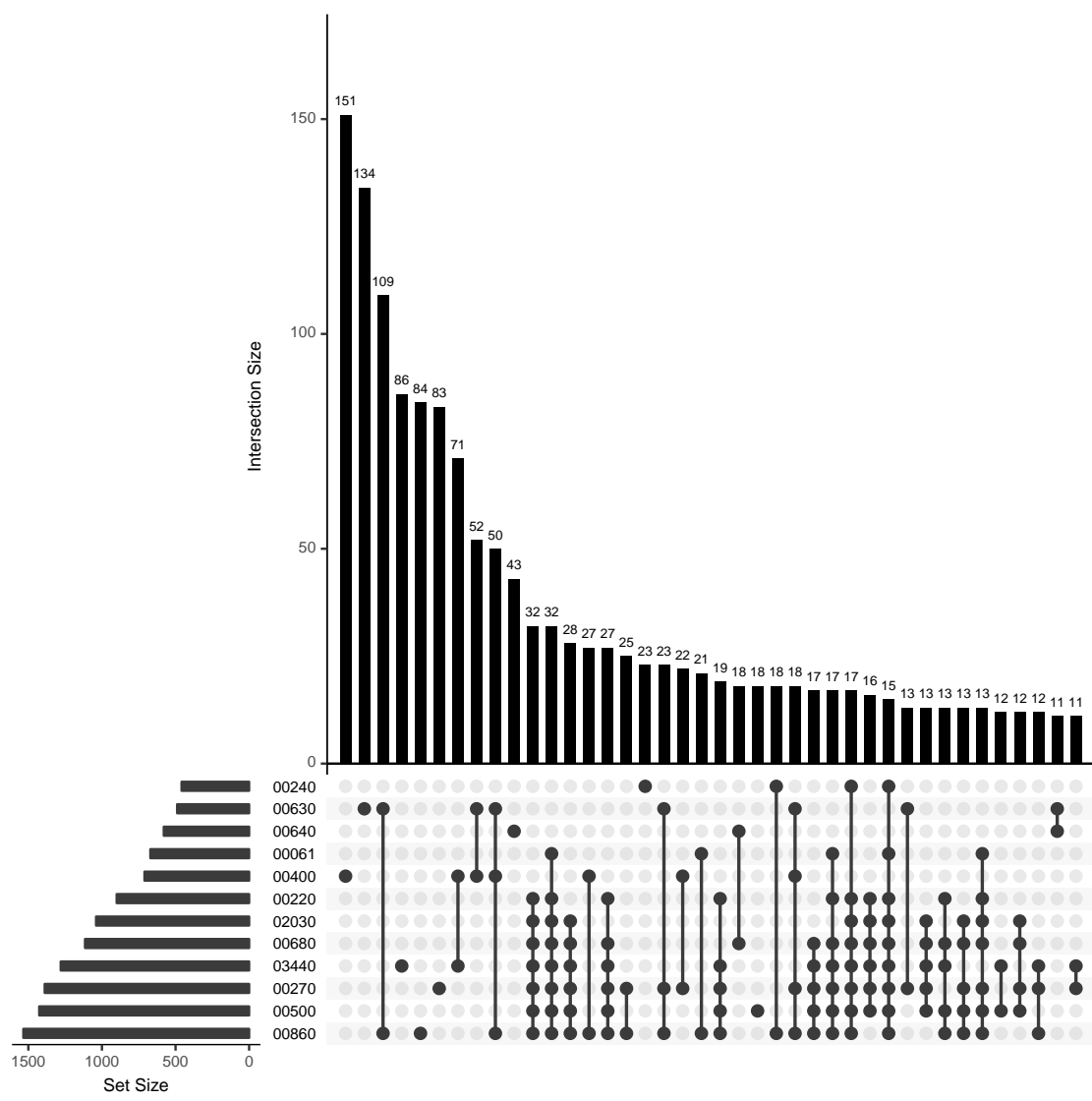


Figure 5: KO LEVEL: KEGG orthology level 1 (pathways) MOC: CAI + mRNA CCs

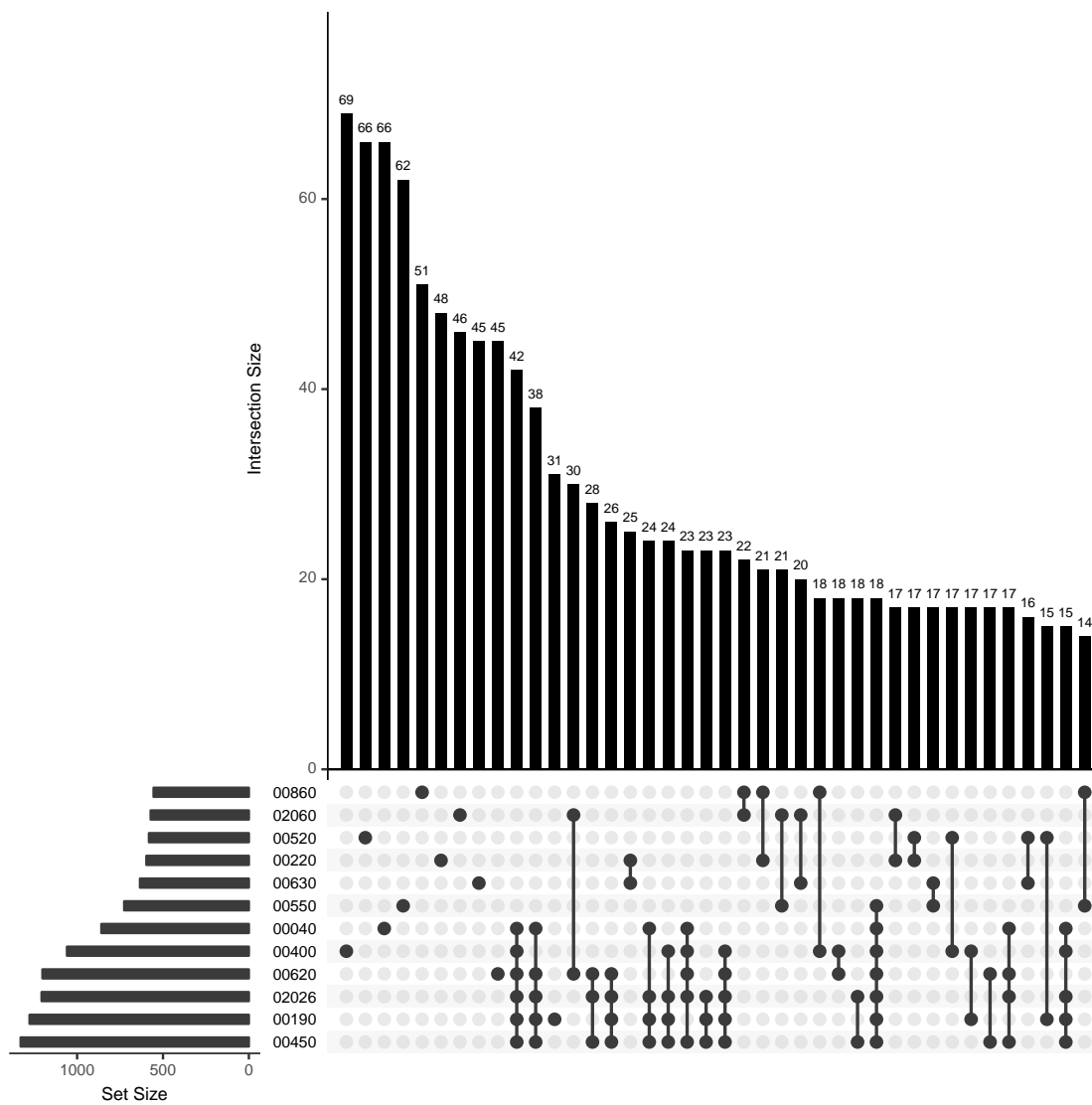


Figure 6: KO LEVEL: KEGG orthology level 1 (pathways) MOC: Molecular Weight + mRNA CCs

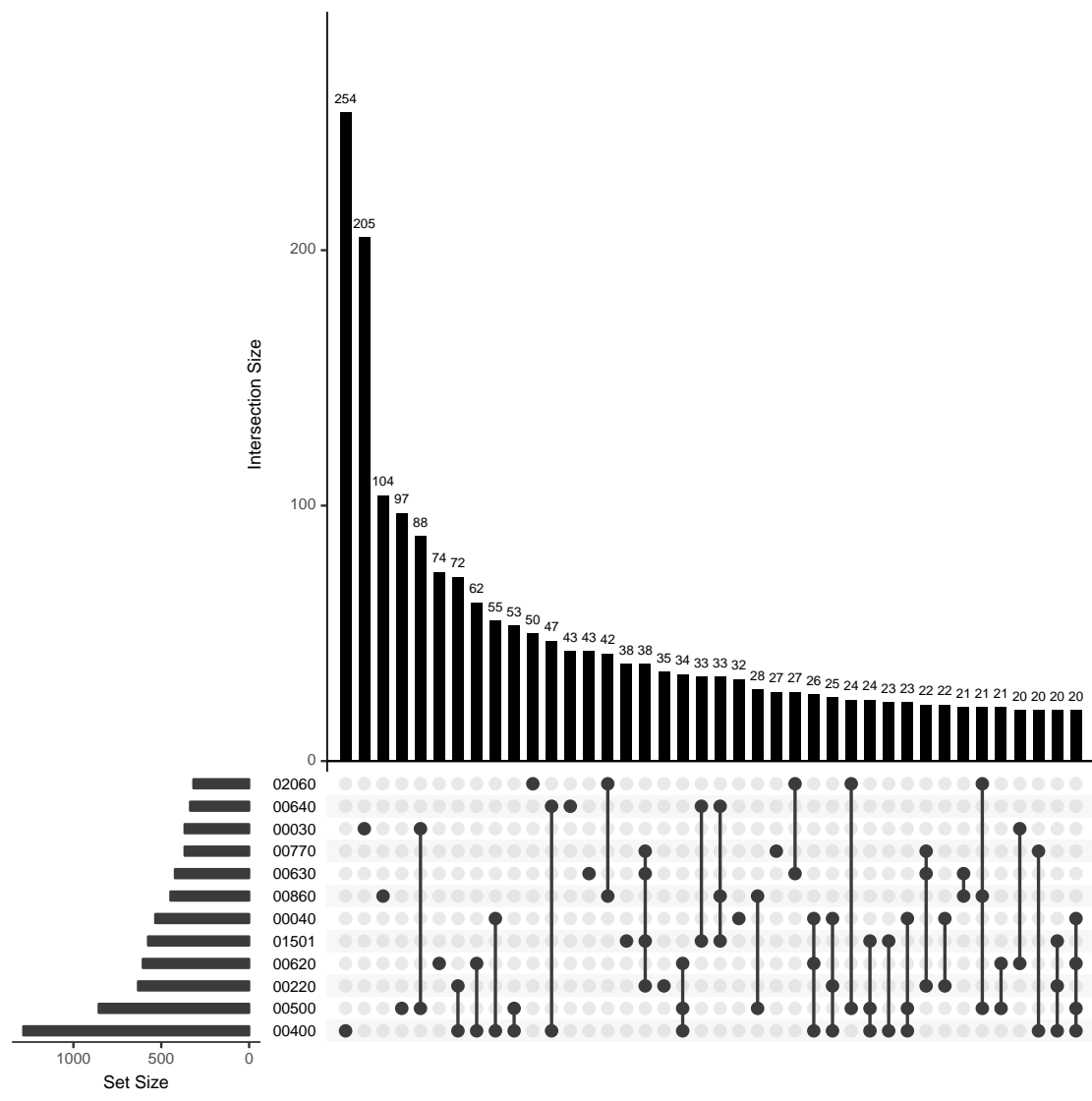


Figure 7: KO LEVEL: KEGG orthology level 1 (pathways) MOC: CAI + Molecular Weight + mRNA CCs

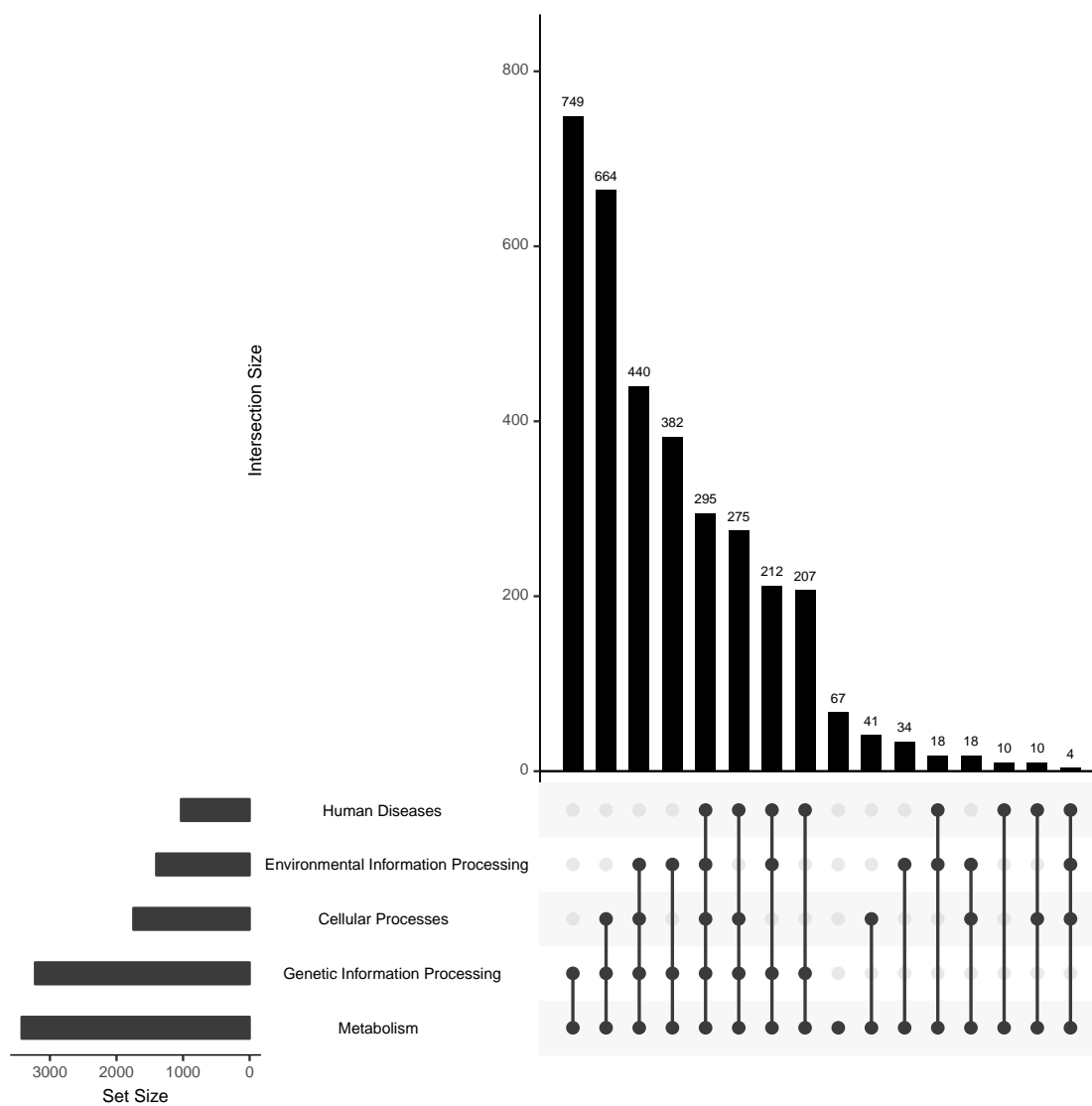


Figure 8: KO LEVEL: KEGG orthology level 2 (functionalities) MOC: CAI + mRNA CCs

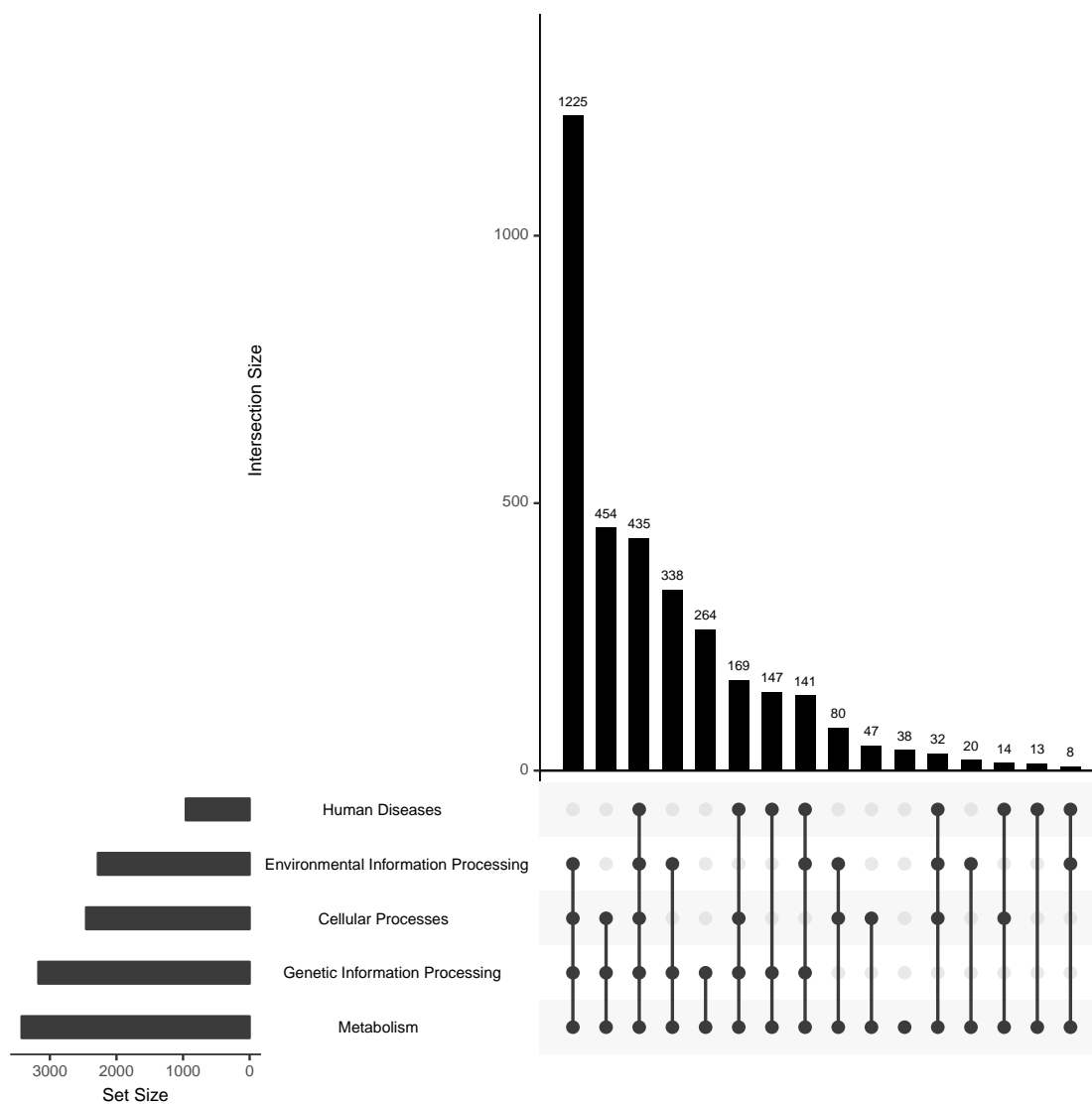


Figure 9: KO LEVEL: KEGG orthology level 2 (functionalities) MOC: Molecular Weight + mRNA CCs

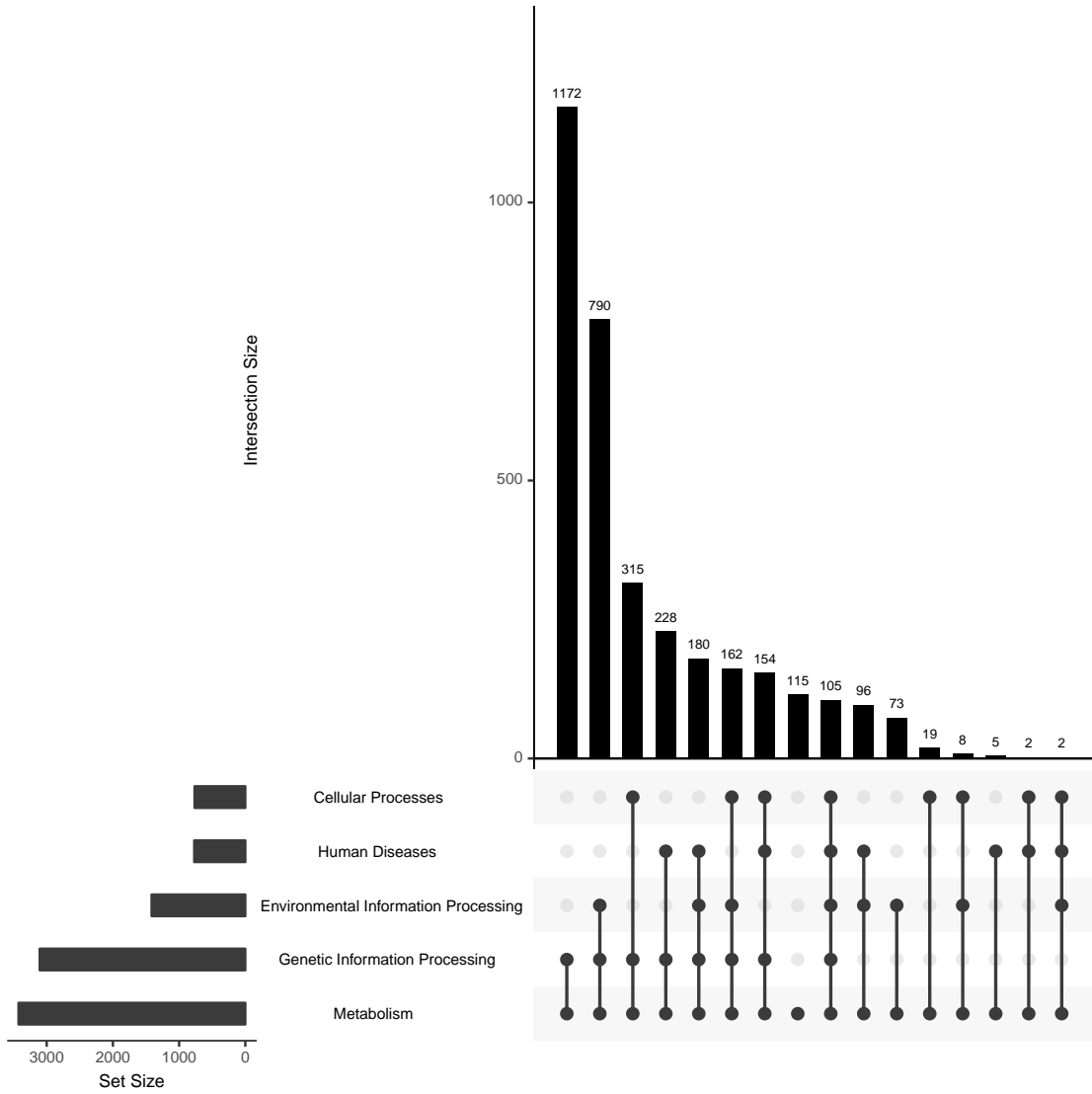


Figure 10: KO LEVEL: KEGG orthology level 2 (functionalities) MOC: CAI + Molecular Weigth + mRNA CCs

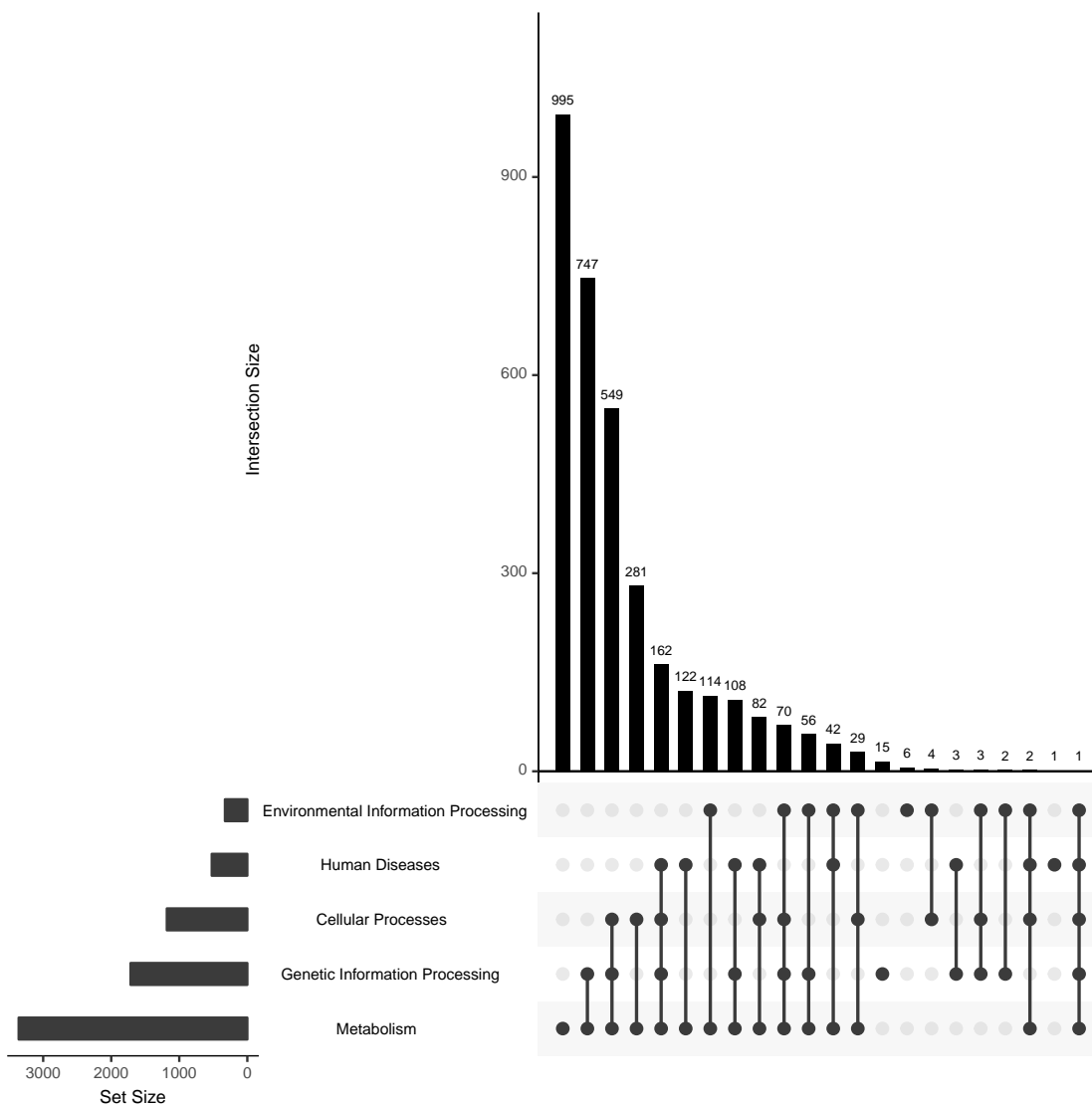


Figure 11: KO LEVEL: KEGG orthology level 2 (functionalities) MOC: CAI + mRNA CCs





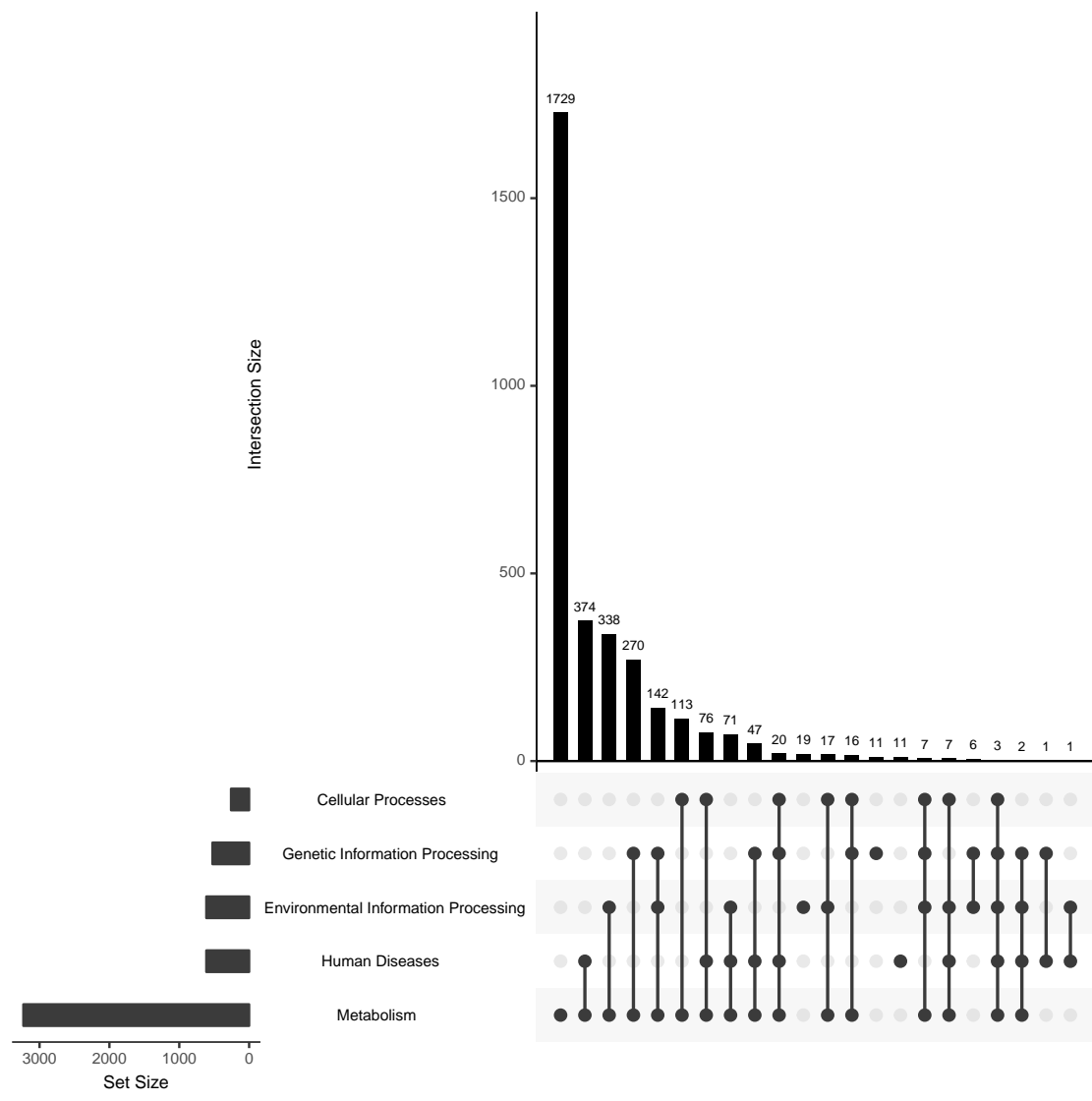


Figure 13: KO LEVEL: KEGG orthology level 2 (functionalities) MOC: CAI + Molecular Weigth + mRNA CCs



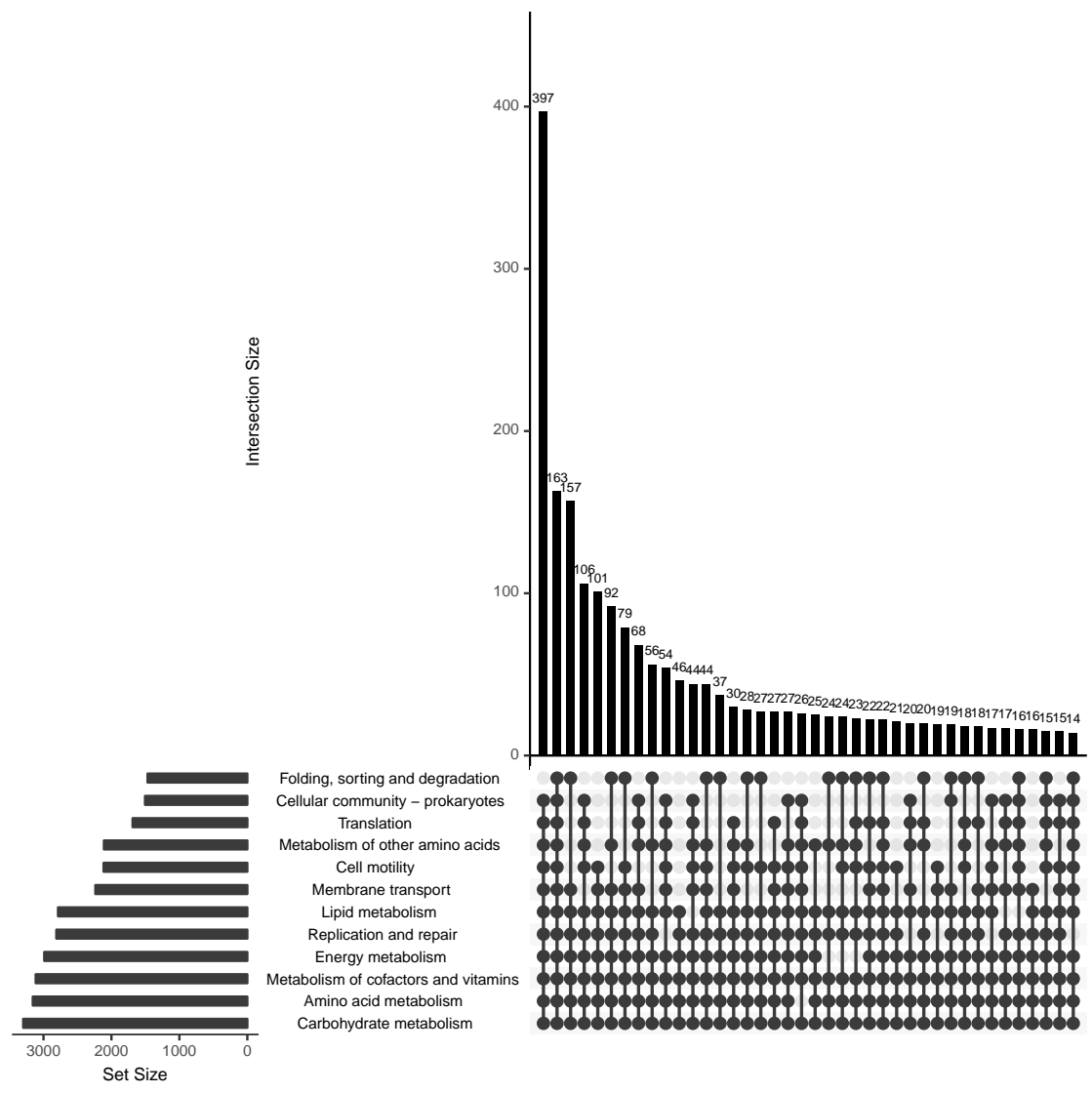


Figure 15: KO LEVEL: KEGG orthology level 3 (pathway maps) MOC: Molecular Weight + mRNA CCs



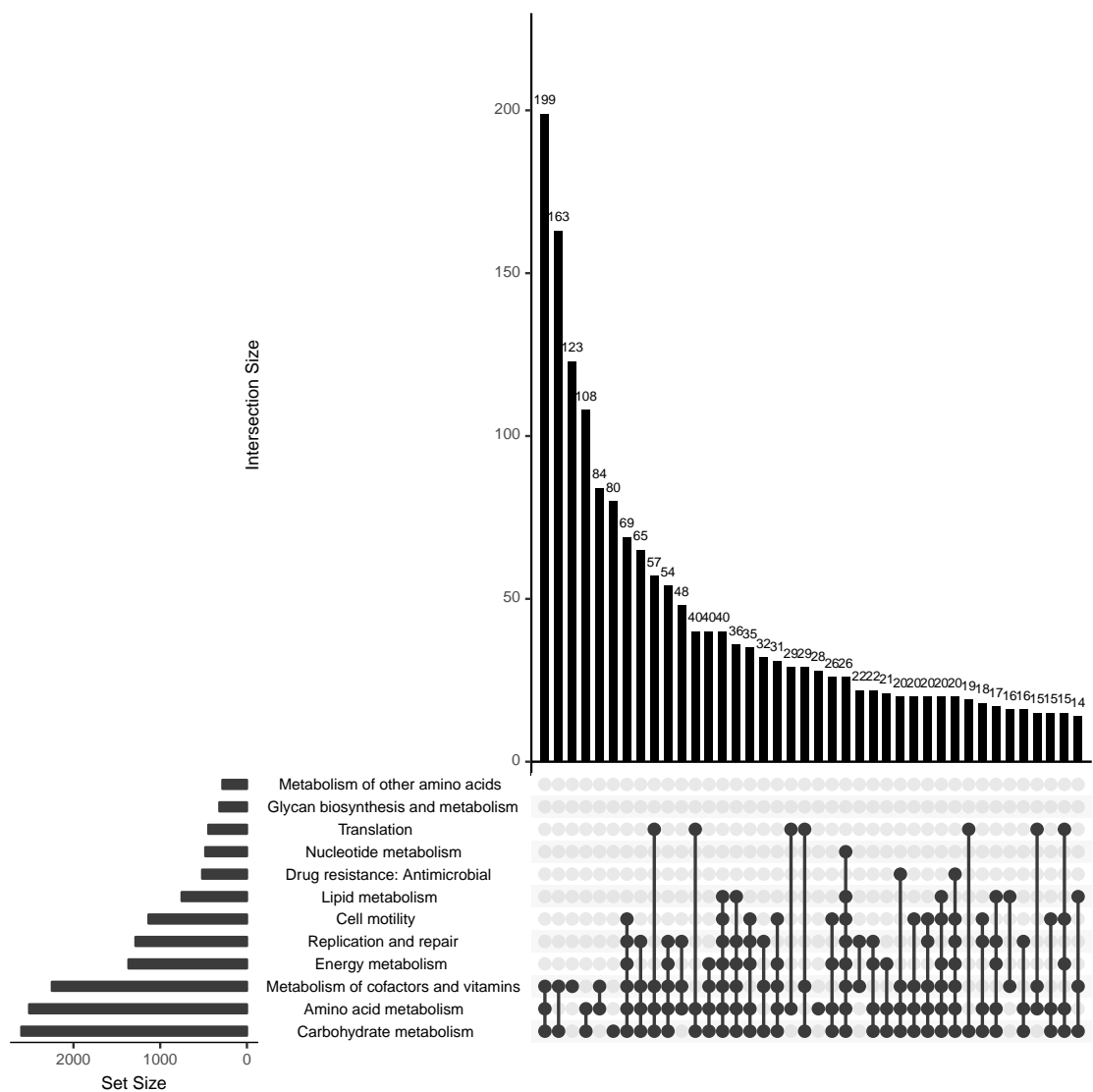


Figure 17: KO LEVEL: KEGG orthology level 3 (pathway maps) MOC: CAI + mRNA CCs

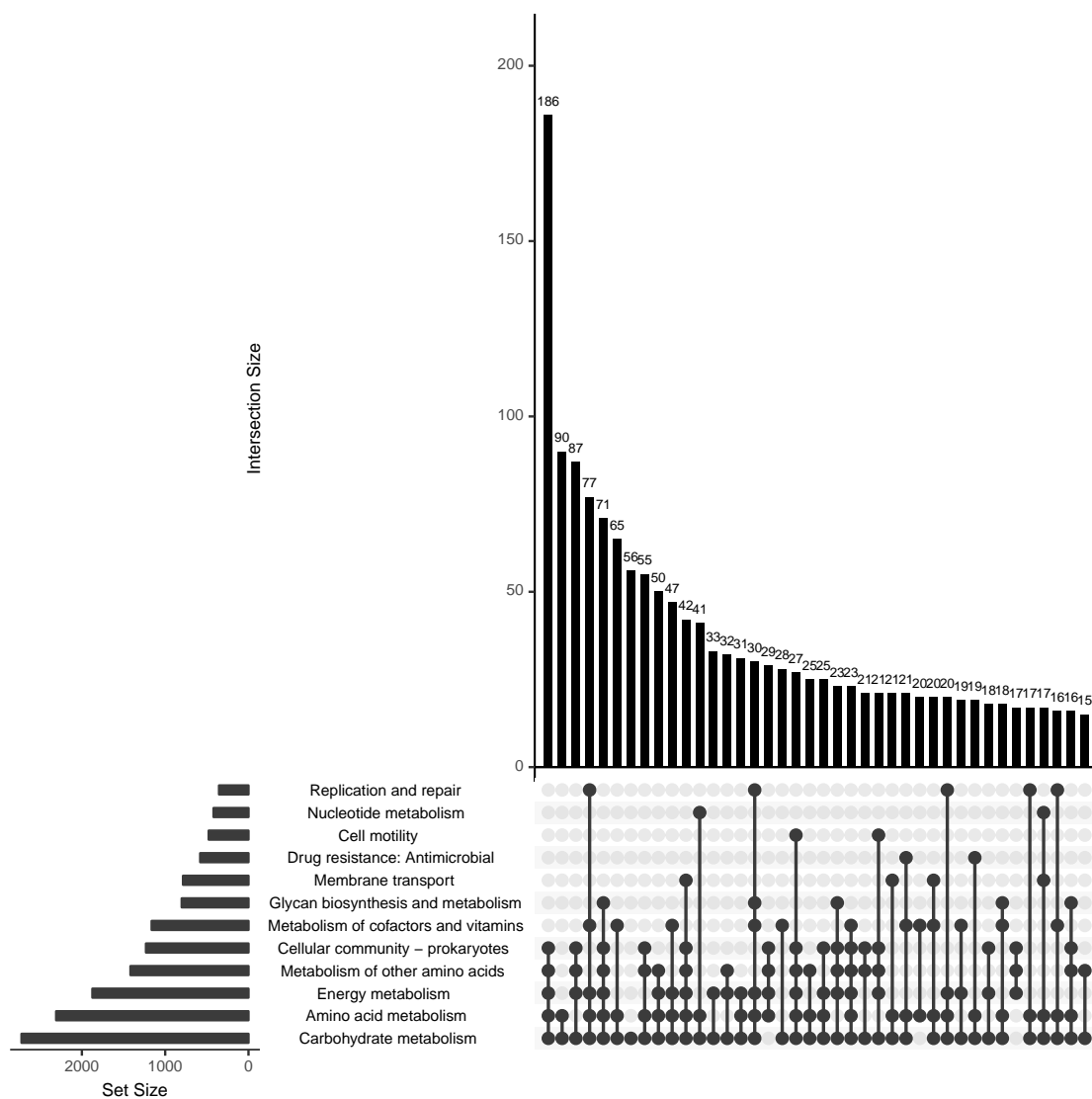


Figure 18: KO LEVEL: KEGG orthology level 3 (pathway maps) MOC: Molecular Weight + mRNA CCs

