Functional profiling from *Babelomics 5*

Dan Crespo

dcrespo@cipf.es

September 29th, 2016

GDA

International Course on

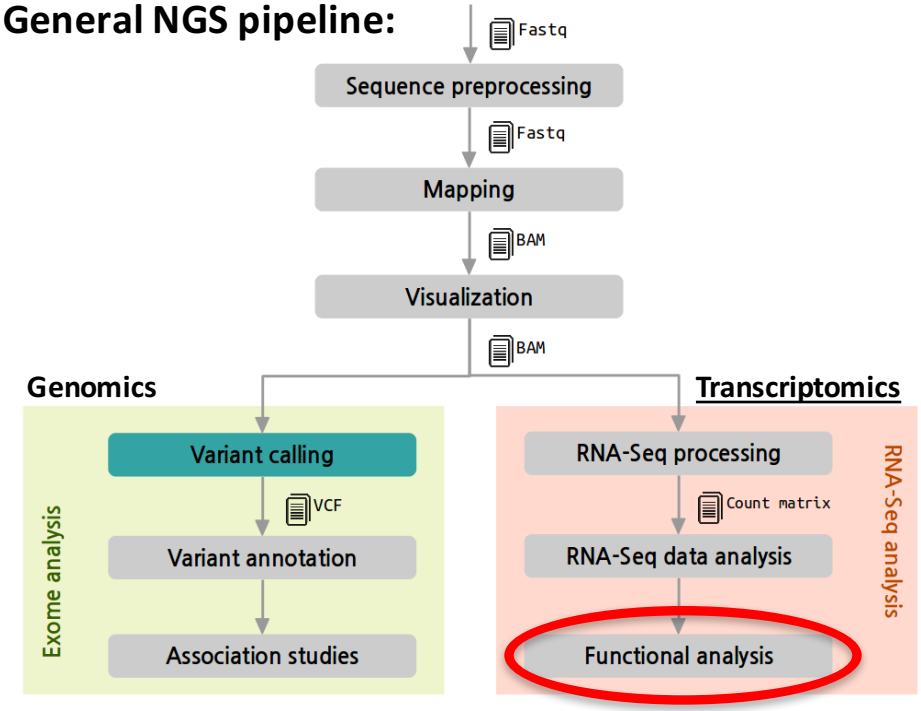
Genomic Data Analysis

PRINCIPE FELIPE

CENTRO DE INVESTIGACION

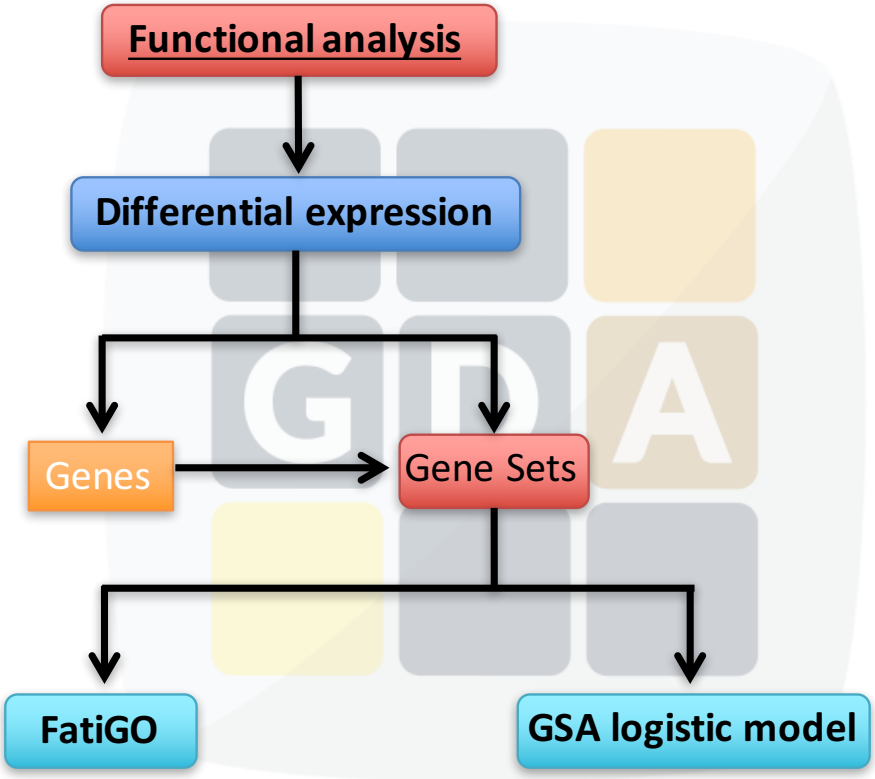
Where are we?

General NGS pipeline:



Functional analysis:

- Single enrichment: **FatiGO**
- Gene set enrichment: **Logistic Model**



Questions we try to answer

- Is there any significant functional enrichment in my gene list / gene sets?

Questions we try to answer

- Is there any significat functional enrichment in my gene list / gene sets?
- Are these genes involved in common pathways?

Questions we try to answer

- Is there any significat functional enrichment in my gene list / gene sets?
- Are these genes involved in common pathways?
- Do they share specific regulation?

Questions we try to answer

- Is there any significat functional enrichment in my gene list / gene sets?
- Are these genes involved in common pathways?
- Do they share specific regulation?
- Are they involved in the same disease?

Babelomics 5: FatiGO

GDA

International Course on

Genomic Data Analysis

PRINCIPE FELIPE

CENTRO DE INVESTIGACION

Single enrichment

Comparison between two list of genes with user defined annotations

Single enrichment

Comparison between two list of genes with user defined annotations

List 1: User provided

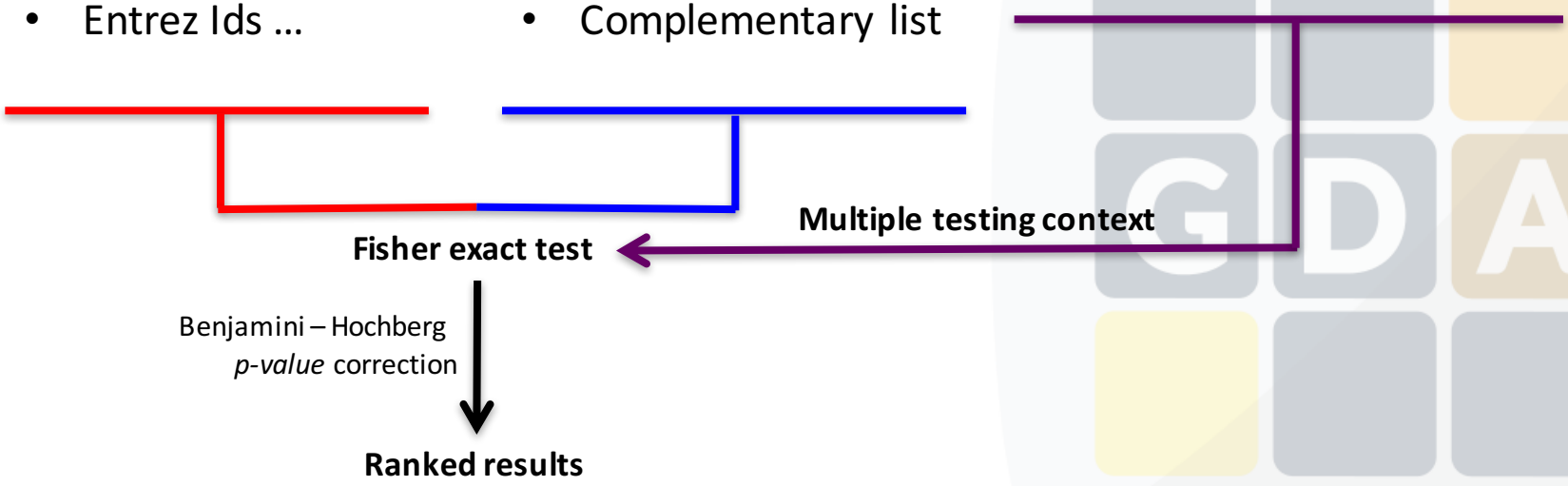
- Gene symbols
- Probe Ids
- Entrez Ids …

List 2:

- User provided
- Rest of genome
- Complementary list

Functional annotations:

- Pre-defined annotations
- Custom annotations



Single enrichment

List 1

AATF	BIRC7	CIAS1	PEA15
APAF1	BIRC8	CIAS1	PNUTL2
API5	BNIP1	CIDEA	PRKAA1
AVEN	BNIP2	CIDEA	RIPK2
BAG1	BNIP3	CIDEB	RTN4
BAG2	BOK	CRADD	SON
BAG3	BRCA1	CRADD	SPHK2
BAG4	CARD10	DAD1	SPP1
BAG5	CARD11	DEDD	SPP1
BAK1	CARD12	DFFA	TNFAIP3
BAX	CARD12	DFFB	TNFRSF10A
BCL10	CARD14	EBAG9	TNFRSF10B
BCL2	CARD15	EBAG9	TNFRSF25
BCL2	CARD4	FADD	TNFRSF6
BCL2A1	CARD4	FAIM2	TNFRSF6B
BCL2L1	CARD6	HDAC1	
BCL2L10	CARD9	HDAC3	
BCL2L13	CASP1	HRK	
BCL2L2	CASP10	HTATIP2	
BCL2L7P1	CASP10	IER3	
BFAR	CASP10	IER3	
BID	CASP2	IER3	
BIK	CASP2	IL10	
BIRC1	CASP2	IL18	
BIRC1	CASP4	IL2	
BIRC2	CASP5	MALT1	
BIRC3	CASP8	MCL1	
BIRC4	CASP9	NME6	
BIRC5	CBX4	NOL3	
BIRC6	CFLAR	PDCD2	

List 2

ACTA1	DNAI1	KIF3B	MYO15A
ACTA2	DNAI2	KIF3C	MYO15B
ACTB	DNAI2	KIF4A	MYO18B
ACTC	DNAL4	KIF5A	MYO1A
ACTG1	DNALI1	KIF5B	MYO1A
ACTG2	DNCH1	KIF5B	MYO1B
ACTL6	DNCI1	KIF5C	MYO1C
ACTR1A	DNCI2	KIF9	MYO1D
ACTR1B	DNCL1	KIFC1	MYO1E
APPBP2	DNCL2A	KIFC3	MYO1E
ATSV	DNCL2B	KNS2	MYO1F
C20orf23	DNCLI2	KNSL7	MYO1G
CENPE	DNM1	MPHOSPH1	MYO3A
DCTN1	DNM2	MYH1	MYO3B
DCTN2	KIF11	MYH1	MYO3B
DNAH1	KIF13A	MYH10	MYO5A
DNAH11	KIF13B	MYH11	MYO5B
DNAH12	KIF14	MYH13	MYO5C
DNAH14	KIF17	MYH13	MYO6
DNAH14	KIF1B	MYH2	MYO6
DNAH14	KIF1C	MYH2	MYO7A
DNAH14	KIF2	MYH3	MYO7B
DNAH14	KIF20A	MYH4	MYO9A
DNAH17	KIF22	MYH4	MYO9B
DNAH2	KIF22	MYH6	OPA1
DNAH3	KIF23	MYH6	RPLP0P1
DNAH5	KIF23	MYH7	RPLP0P1
DNAH7	KIF25	MYH7B	TCTE1
DNAH8	KIF26A	MYH8	TCTEL1
DNAH9	KIF2C	MYH9	TIAF1
DNAH9	KIF3A	MYO10	

Functional annotations

Apoptotic process GO:0006915

CBX4	BRCA1	AVEN	MCL1
CRADD	HDAC3	FAIM2	BCL2L2
IL10	HDAC1	BID	BCL2L1
BAK1	RTN4	TNFRSF25	PEA15
BIRC8	IER3	BCL2	CFLAR
API5	HTATIP2	DAD1	CARD9
BCL10	DEDD	PRKAA1	SPHK2
OPA1	BNIP3	ACTC	CIDEB
DFFA	PDCD2	BIRC6	CASP2
DFFB	CASP5	BIRC5	EBAG9
BCL2A1	BAG5	BIRC3	NOL3
HRK	BAG4	BIRC2	CARD10
FADD	CASP4	BIRC1	TNFRSF10A
BCL2L13	CASP9	CASP10	CARD11
BCL2L10	BAG1	SON	CARD14
NME6	BOK	BAX	KIF1B
RIPK2	BAG3	BIK	DNM2
TNFAIP3	CASP8	APAF1	TNFRSF10B
TNFRSF6B	CASP1	TIAF1	BNIP1
BNIP2	CARD6	MALT1	CIDEA

Single enrichment

List 1

AATF	BIRC7	CIAS1	PEA15
APAF1	BIRC8	CIAS1	PNUTL2
API5	BNIP1	CIDEA	PRKAA1
AVEN	BNIP2	CIDEA	RIPK2
BAG1	BNIP3	CIDEB	RTN4
BAG2	BOK	CRADD	SON
BAG3	BRCA1	CRADD	SPHK2
BAG4	CARD10	DAD1	SPP1
BAG5	CARD11	DEDD	SPP1
BAK1	CARD12	DFFA	TNFAIP3
BAX	CARD12	DFFB	TNFRSF10A
BCL10	CARD14	EBAG9	TNFRSF10B
BCL2	CARD15	EBAG9	TNFRSF25
BCL2	CARD4	FADD	TNFRSF6
BCL2A1	CARD4	FAIM2	TNFRSF6B
BCL2L1	CARD6	HDAC1	
BCL2L10	CARD9	HDAC3	
BCL2L13	CASP1	HRK	
BCL2L2	CASP10	HTATIP2	
BCL2L7P1	CASP10	IER3	
BFAR	CASP10	IER3	
BID	CASP2	IER3	
BIK	CASP2	IL10	
BIRC1	CASP2	IL18	
BIRC1	CASP4	IL2	
BIRC2	CASP5	MALT1	
BIRC3	CASP8	MCL1	
BIRC4	CASP9	NME6	
BIRC5	CBX4	NOL3	
BIRC6	CFLAR	PDCD2	

List 2

ACTA1	DNAI1	KIF3B	MYO15A
ACTA2	DNAI2	KIF3C	MYO15B
ACTB	DNAI2	KIF4A	MYO18B
ACTC	DNAL4	KIF5A	MYO1A
ACTG1	DNALI1	KIF5B	MYO1A
ACTG2	DNCH1	KIF5B	MYO1B
ACTL6	DNCI1	KIF5C	MYO1C
ACTR1A	DNCI2	KIF9	MYO1D
ACTR1B	DNCL1	KIFC1	MYO1E
APPBP2	DNCL2A	KIFC3	MYO1E
ATSV	DNCL2B	KNS2	MYO1F
C20orf23	DNCLI2	KNSL7	MYO1G
CENPE	DNM1	MPHOSPH1	MYO3A
DCTN1	DNM2	MYH1	MYO3B
DCTN2	KIF11	MYH1	MYO3B
DNAH1	KIF13A	MYH10	MYO5A
DNAH11	KIF13B	MYH11	MYO5B
DNAH12	KIF14	MYH13	MYO5C
DNAH14	KIF17	MYH13	MYO6
DNAH14	KIF1B	MYH2	MYO6
DNAH14	KIF1C	MYH2	MYO7A
DNAH14	KIF2	MYH3	MYO7B
DNAH14	KIF20A	MYH4	MYO9A
DNAH17	KIF22	MYH4	MYO9B
DNAH2	KIF22	MYH6	OPA1
DNAH3	KIF23	MYH6	RPLP0P1
DNAH5	KIF23	MYH7	RPLP0P1
DNAH7	KIF25	MYH7B	TCTE1
DNAH8	KIF26A	MYH8	TCTEL1
DNAH9	KIF2C	MYH9	TIAF1
DNAH9	KIF3A	MYO10	

Functional annotations

Apoptotic process GO:0006915

CBX4	BRCA1	AVEN	MCL1
CRADD	HDAC3	FAIM2	BCL2L2
IL10	HDAC1	BID	BCL2L1
BAK1	RTN4	TNFRSF25	PEA15
BIRC8	IER3	BCL2	CFLAR
API5	HTATIP2	DAD1	CARD9
BCL10	DEDD	PRKAA1	SPHK2
OPA1	BNIP3	ACTC	CIDEB
DFFA	PDCD2	BIRC6	CASP2
DFFB	CASP5	BIRC5	EBAG9
BCL2A1	BAG5	BIRC3	NOL3
HRK	BAG4	BIRC2	CARD10
FADD	CASP4	BIRC1	TNFRSF10A
BCL2L13	CASP9	CASP10	CARD11
BCL2L10	BAG1	SON	CARD14
NME6	BOK	BAX	KIF1B
RIPK2	BAG3	BIK	DNM2
TNFAIP3	CASP8	APAF1	TNFRSF10B
TNFRSF6B	CASP1	TIAF1	BNIP1
BNIP2	CARD6	MALT1	CIDEA

Single enrichment

List 1

AATF	BIRC7	CIAS1	PEA15
APAF1	BIRC8	CIAS1	PNUTL2
API5	BNIP1	CIDEA	PRKAA1
AVEN	BNIP2	CIDEA	RIPK2
BAG1	BNIP3	CIDEB	RTN4
BAG2	BOK	CRADD	SON
BAG3	BRCA1	CRADD	SPHK2
BAG4	CARD10	DAD1	SPP1
BAG5	CARD11	DEDD	SPP1
BAK1	CARD12	DFFA	TNFAIP3
BAX	CARD12	DFFB	TNFRSF10A
BCL10	CARD14	EBAG9	TNFRSF10B
BCL2	CARD15	EBAG9	TNFRSF25
BCL2	CARD4	FADD	TNFRSF6
BCL2A1	CARD4	FAIM2	TNFRSF6B
BCL2L1	CARD6	HDAC1	
BCL2L10	CARD9	HDAC3	
BCL2L13	CASP1	HRK	
BCL2L2	CASP10	HTATIP2	
BCL2L7P1	CASP10	IER3	
BFAR	CASP10	IER3	
BID	CASP2	IER3	
BIK	CASP2	IL10	
BIRC1	CASP2	IL18	
BIRC1	CASP4	IL2	
BIRC2	CASP5	MALT1	
BIRC3	CASP8	MCL1	
BIRC4	CASP9	NME6	
BIRC5	CBX4	NOL3	
BIRC6	CFLAR	PDCD2	

List 2

ACTA1	DNAI1	KIF3B	MYO15A
ACTA2	DNAI2	KIF3C	MYO15B
ACTB	DNAI2	KIF4A	MYO18B
ACTC	DNAL4	KIF5A	MYO1A
ACTG1	DNALI1	KIF5B	MYO1A
ACTG2	DNCH1	KIF5B	MYO1B
ACTL6	DNCI1	KIF5C	MYO1C
ACTR1A	DNCI2	KIF9	MYO1D
ACTR1B	DNCL1	KIFC1	MYO1E
APPBP2	DNCL2A	KIFC3	MYO1E
ATSV	DNCL2B	KNS2	MYO1F
C20orf23	DNCLI2	KNSL7	MYO1G
CENPE	DNM1	MPHOSPH1	MYO3A
DCTN1	DNM2	MYH1	MYO3B
DCTN2	KIF11	MYH1	MYO3B
DNAH1	KIF13A	MYH10	MYO5A
DNAH11	KIF13B	MYH11	MYO5B
DNAH12	KIF14	MYH13	MYO5C
DNAH14	KIF17	MYH13	MYO6
DNAH14	KIF1B	MYH2	MYO6
DNAH14	KIF1C	MYH2	MYO7A
DNAH14	KIF2	MYH3	MYO7B
DNAH14	KIF20A	MYH4	MYO9A
DNAH17	KIF22	MYH4	MYO9B
DNAH2	KIF22	MYH6	OPA1
DNAH3	KIF23	MYH6	RPLP0P1
DNAH5	KIF23	MYH7	RPLP0P1
DNAH7	KIF25	MYH7B	TCTE1
DNAH8	KIF26A	MYH8	TCTEL1
DNAH9	KIF2C	MYH9	TIAF1
DNAH9	KIF3A	MYO10	

Functional annotations

Apoptotic process GO:0006915

CBX4	BRCA1	AVEN	MCL1
CRADD	HDAC3	FAIM2	BCL2L2
IL10	HDAC1	BID	BCL2L1
BAK1	RTN4	TNFRSF25	PEA15
BIRC8	IER3	BCL2	CFLAR
API5	HTATIP2	DAD1	CARD9
BCL10	DEDD	PRKAA1	SPHK2
OPA1	BNIP3	ACTC	CIDEB
DFFA	PDCD2	BIRC6	CASP2
DFFB	CASP5	BIRC5	EBAG9
BCL2A1	BAG5	BIRC3	NOL3
HRK	BAG4	BIRC2	CARD10
FADD	CASP4	BIRC1	TNFRSF10A
BCL2L13	CASP9	CASP10	CARD11
BCL2L10	BAG1	SON	CARD14
NME6	BOK	BAX	KIF1B
RIPK2	BAG3	BIK	DNM2
TNFAIP3	CASP8	APAF1	TNFRSF10B
TNFRSF6B	CASP1	TIAF1	BNIP1
BNIP2	CARD6	MALT1	CIDEA

Single enrichment

List 1

AATF	BIRC7	CIAS1	PEA15
APAF1	BIRC8	CIAS1	PNUTL2
API5	BNIP1	CIDEA	PRKAA1
AVEN	BNIP2	CIDEA	RIPK2
BAG1	BNIP3	CIDEB	RTN4
BAG2	BOK	CRADD	SON
BAG3	BRCA1	CRADD	SPHK2
BAG4	CARD10	DAD1	SPP1
BAG5	CARD11	DEDD	SPP1
BAK1	CARD12	DFFA	TNFAIP3
BAX	CARD12	DFFB	TNFRSF10A
BCL10	CARD14	EBAG9	TNFRSF10B
BCL2	CARD15	EBAG9	TNFRSF25
BCL2	CARD4	FADD	TNFRSF6
BCL2A1	CARD4	FAIM2	TNFRSF6B
BCL2L1	CARD6	HDAC1	
BCL2L10	CARD9	HDAC3	
BCL2L13	CASP1	HRK	
BCL2L2	CASP10	HTATIP2	
BCL2L7P1	CASP10	IER3	
BFAR	CASP10	IER3	
BID	CASP2	IER3	
BIK	CASP2	IL10	
BIRC1	CASP2	IL18	
BIRC1	CASP4	IL2	
BIRC2	CASP5	MALT1	
BIRC3	CASP8	MCL1	
BIRC4	CASP9	NME6	
BIRC5	CBX4	NOL3	
BIRC6	CFLAR	PDCD2	

List 2

ACTA1	DNAI1	KIF3B	MYO15A
ACTA2	DNAI2	KIF3C	MYO15B
ACTB	DNAI2	KIF4A	MYO18B
ACTC	DNAL4	KIF5A	MYO1A
ACTG1	DNALI1	KIF5B	MYO1A
ACTG2	DNCH1	KIF5B	MYO1B
ACTL6	DNCI1	KIF5C	MYO1C
ACTR1A	DNCI2	KIF9	MYO1D
ACTR1B	DNCL1	KIFC1	MYO1E
APPBP2	DNCL2A	KIFC3	MYO1E
ATSV	DNCL2B	KNS2	MYO1F
C20orf23	DNCLI2	KNSL7	MYO1G
CENPE	DNM1	MPHOSPH1	MYO3A
DCTN1	DNM2	MYH1	MYO3B
DCTN2	KIF11	MYH1	MYO3B
DNAH1	KIF13A	MYH10	MYO5A
DNAH11	KIF13B	MYH11	MYO5B
DNAH12	KIF14	MYH13	MYO5C
DNAH14	KIF17	MYH13	MYO6
DNAH14	KIF1B	MYH2	MYO6
DNAH14	KIF1C	MYH2	MYO7A
DNAH14	KIF2	MYH3	MYO7B
DNAH14	KIF20A	MYH4	MYO9A
DNAH17	KIF22	MYH4	MYO9B
DNAH2	KIF22	MYH6	OPA1
DNAH3	KIF23	MYH6	OPA1
DNAH5	KIF23	MYH7	RPLP0P1
DNAH7	KIF25	MYH7B	RPLP0P1
DNAH8	KIF26A	MYH8	TCTE1
DNAH9	KIF2C	MYH9	TCTEL1
DNAH9	KIF3A	MYO10	TIAF1

Functional annotations

Positive regulation of immune system process GO0002684

IL18	RIPK2	MALT1	ACTB
ACTG1	TNFAIP3	MYO10	BIRC3
BCL10	CARD9	CASP8	BIRC2
MYH2	MYO1C	CARD11	BAX
FADD	MYO1G	BCL2	IL2

Single enrichment

List 1

AATF	BIRC7	CIAS1	PEA15
APAF1	BIRC8	CIAS1	PNUTL2
API5	BNIP1	CIDEA	PRKAA1
AVEN	BNIP2	CIDEA	RIPK2
BAG1	BNIP3	CIDEB	RTN4
BAG2	BOK	CRADD	SON
BAG3	BRCA1	CRADD	SPHK2
BAG4	CARD10	DAD1	SPP1
BAG5	CARD11	DEDD	SPP1
BAK1	CARD12	DFFA	TNFAIP3
BAX	CARD12	DFFB	TNFRSF10A
BCL10	CARD14	EBAG9	TNFRSF10B
BCL2	CARD15	EBAG9	TNFRSF25
BCL2	CARD4	FADD	TNFRSF6
BCL2A1	CARD4	FAIM2	TNFRSF6B
BCL2L1	CARD6	HDAC1	
BCL2L10	CARD9	HDAC3	
BCL2L13	CASP1	HRK	
BCL2L2	CASP10	HTATIP2	
BCL2L7P1	CASP10	IER3	
BFAR	CASP10	IER3	
BID	CASP2	IER3	
BIK	CASP2	IL10	
BIRC1	CASP2	IL18	
BIRC1	CASP4	IL2	
BIRC2	CASP5	MALT1	
BIRC3	CASP8	MCL1	
BIRC4	CASP9	NME6	
BIRC5	CBX4	NOL3	
BIRC6	CFLAR	PDCD2	

List 2

ACTA1	DNAI1	KIF3B	MYO15A
ACTA2	DNAI2	KIF3C	MYO15B
ACTB	DNAI2	KIF4A	MYO18B
ACTC	DNAL4	KIF5A	MYO1A
ACTG1	DNALI1	KIF5B	MYO1A
ACTG2	DNCH1	KIF5B	MYO1B
ACTL6	DNCI1	KIF5C	MYO1C
ACTR1A	DNCI2	KIF9	MYO1D
ACTR1B	DNCL1	KIFC1	MYO1E
APPBP2	DNCL2A	KIFC3	MYO1E
ATSV	DNCL2B	KNS2	MYO1F
C20orf23	DNCLI2	KNSL7	MYO1G
CENPE	DNM1	MPHOSPH1	MYO3A
DCTN1	DNM2	MYH1	MYO3B
DCTN2	KIF11	MYH1	MYO3B
DNAH1	KIF13A	MYH10	MYO5A
DNAH11	KIF13B	MYH11	MYO5B
DNAH12	KIF14	MYH13	MYO5C
DNAH14	KIF17	MYH13	MYO6
DNAH14	KIF1B	MYH2	MYO6
DNAH14	KIF1C	MYH2	MYO7A
DNAH14	KIF2	MYH3	MYO7B
DNAH14	KIF20A	MYH4	MYO9A
DNAH17	KIF22	MYH4	MYO9B
DNAH2	KIF22	MYH6	OPA1
DNAH3	KIF23	MYH6	OPA1
DNAH5	KIF23	MYH7	RPLP0P1
DNAH7	KIF25	MYH7B	RPLP0P1
DNAH8	KIF26A	MYH8	TCTE1
DNAH9	KIF2C	MYH9	TCTEL1
DNAH9	KIF3A	MYO10	TIAF1

Functional annotations

Positive regulation of immune system process GO0002684

IL18	RIPK2	MALT1	ACTB
ACTG1	TNFAIP3	MYO10	BIRC3
BCL10	CARD9	CASP8	BIRC2
MYH2	MYO1C	CARD11	BAX
FADD	MYO1G	BCL2	IL2

Single enrichment

List 1

AATF	BIRC7	CIAS1	PEA15
APAF1	BIRC8	CIAS1	PNUTL2
API5	BNIP1	CIDEA	PRKAA1
AVEN	BNIP2	CIDEA	RIPK2
BAG1	BNIP3	CIDEB	RTN4
BAG2	BOK	CRADD	SON
BAG3	BRCA1	CRADD	SPHK2
BAG4	CARD10	DAD1	SPP1
BAG5	CARD11	DEDD	SPP1
BAK1	CARD12	DFFA	TNFAIP3
BAX	CARD12	DFFB	TNFRSF10A
BCL10	CARD14	EBAG9	TNFRSF10B
BCL2	CARD15	EBAG9	TNFRSF25
BCL2	CARD4	FADD	TNFRSF6
BCL2A1	CARD4	FAIM2	TNFRSF6B
BCL2L1	CARD6	HDAC1	
BCL2L10	CARD9	HDAC3	
BCL2L13	CASP1	HRK	
BCL2L2	CASP10	HTATIP2	
BCL2L7P1	CASP10	IER3	
BFAR	CASP10	IER3	
BID	CASP2	IER3	
BIK	CASP2	IL10	
BIRC1	CASP2	IL18	
BIRC1	CASP4	IL2	
BIRC2	CASP5	MALT1	
BIRC3	CASP8	MCL1	
BIRC4	CASP9	NME6	
BIRC5	CBX4	NOL3	
BIRC6	CFLAR	PDCD2	

List 2

ACTA1	DNAI1	KIF3B	MYO15A
ACTA2	DNAI2	KIF3C	MYO15B
ACTB	DNAI2	KIF4A	MYO18B
ACTC	DNAL4	KIF5A	MYO1A
ACTG1	DNALI1	KIF5B	MYO1A
ACTG2	DNCH1	KIF5B	MYO1B
ACTL6	DNCI1	KIF5C	MYO1C
ACTR1A	DNCI2	KIF9	MYO1D
ACTR1B	DNCL1	KIFC1	MYO1E
APPBP2	DNCL2A	KIFC3	MYO1E
ATSV	DNCL2B	KNS2	MYO1F
C20orf23	DNCLI2	KNSL7	MYO1G
CENPE	DNM1	MPHOSPH1	MYO3A
DCTN1	DNM2	MYH1	MYO3B
DCTN2	KIF11	MYH1	MYO3B
DNAH1	KIF13A	MYH10	MYO5A
DNAH11	KIF13B	MYH11	MYO5B
DNAH12	KIF14	MYH13	MYO5C
DNAH14	KIF17	MYH13	MYO6
DNAH14	KIF1B	MYH2	MYO6
DNAH14	KIF1C	MYH2	MYO7A
DNAH14	KIF2	MYH3	MYO7B
DNAH14	KIF20A	MYH4	MYO9A
DNAH17	KIF22	MYH4	MYO9B
DNAH2	KIF22	MYH6	OPA1
DNAH3	KIF23	MYH6	OPA1
DNAH5	KIF23	MYH7	RPLP0P1
DNAH7	KIF25	MYH7B	RPLP0P1
DNAH8	KIF26A	MYH8	TCTE1
DNAH9	KIF2C	MYH9	TCTEL1
DNAH9	KIF3A	MYO10	TIAF1

Functional annotations

Positive regulation of immune system process GO0002684

IL18	RIPK2	MALT1	ACTB
ACTG1	TNFAIP3	MYO10	BIRC3
BCL10	CARD9	CASP8	BIRC2
MYH2	MYO1C	CARD11	BAX
FADD	MYO1G	BCL2	IL2

Organ development GO0048515

IL18	MYH6	KIF3A	MYO7A
IL10	MYH9	SPHK2	CASP5
BAK1	MYO18B	MYO1E	CASP9
MYO6	RIPK2	MALT1	BOK
MYO3A	BCL2L2	HDAC1	BAG3
MYH3	BCL2L1	MYH11	CASP8
DFFB	MYO15A	MYH10	CASP2
MYH7	SPP1	RTN4	ACTA1
FADD	CFLAR	DEDD	ACTA2
ACTB	DNAH5	MYO5A	CARD11
ACTC	BCL2	BID	FAIM2
BIRC6	BAX	BIK	APAF1
IL2			

Single enrichment

List 1

AATF	BIRC7	CIAS1	PEA15
APAF1	BIRC8	CIAS1	PNUTL2
API5	BNIP1	CIDEA	PRKAA1
AVEN	BNIP2	CIDEA	RIPK2
BAG1	BNIP3	CIDEB	RTN4
BAG2	BOK	CRADD	SON
BAG3	BRCA1	CRADD	SPHK2
BAG4	CARD10	DAD1	SPP1
BAG5	CARD11	DEDD	SPP1
BAK1	CARD12	DFFA	TNFAIP3
BAX	CARD12	DFFB	TNFRSF10A
BCL10	CARD14	EBAG9	TNFRSF10B
BCL2	CARD15	EBAG9	TNFRSF25
BCL2	CARD4	FADD	TNFRSF6
BCL2A1	CARD4	FAIM2	TNFRSF6B
BCL2L1	CARD6	HDAC1	
BCL2L10	CARD9	HDAC3	
BCL2L13	CASP1	HRK	
BCL2L2	CASP10	HTATIP2	
BCL2L7P1	CASP10	IER3	
BFAR	CASP10	IER3	
BID	CASP2	IER3	
BIK	CASP2	IL10	
BIRC1	CASP2	IL18	
BIRC1	CASP4	IL2	
BIRC2	CASP5	MALT1	
BIRC3	CASP8	MCL1	
BIRC4	CASP9	NME6	
BIRC5	CBX4	NOL3	
BIRC6	CFLAR	PDCD2	

List 2

ACTA1	DNAI1	KIF3B	MYO15A
ACTA2	DNAI2	KIF3C	MYO15B
ACTB	DNAI2	KIF4A	MYO18B
ACTC	DNAL4	KIF5A	MYO1A
ACTG1	DNALI1	KIF5B	MYO1A
ACTG2	DNCH1	KIF5B	MYO1B
ACTL6	DNCI1	KIF5C	MYO1C
ACTR1A	DNCI2	KIF9	MYO1D
ACTR1B	DNCL1	KIFC1	MYO1E
APPBP2	DNCL2A	KIFC3	MYO1E
ATSV	DNCL2B	KNS2	MYO1F
C20orf23	DNCLI2	KNSL7	MYO1G
CENPE	DNM1	MPHOSPH1	MYO3A
DCTN1	DNM2	MYH1	MYO3B
DCTN2	KIF11	MYH1	MYO3B
DNAH1	KIF13A	MYH10	MYO5A
DNAH11	KIF13B	MYH11	MYO5B
DNAH12	KIF14	MYH13	MYO5C
DNAH14	KIF17	MYH13	MYO6
DNAH14	KIF1B	MYH2	MYO6
DNAH14	KIF1C	MYH2	MYO7A
DNAH14	KIF2	MYH3	MYO7B
DNAH14	KIF20A	MYH4	MYO9A
DNAH17	KIF22	MYH4	MYO9B
DNAH2	KIF22	MYH6	OPA1
DNAH3	KIF23	MYH6	OPA1
DNAH5	KIF23	MYH7	RPLP0P1
DNAH7	KIF25	MYH7B	RPLP0P1
DNAH8	KIF26A	MYH8	TCTE1
DNAH9	KIF2C	MYH9	TCTEL1
DNAH9	KIF3A	MYO10	TIAF1

Functional annotations

Positive regulation of immune system process GO0002684

IL18	RIPK2	MALT1	ACTB
ACTG1	TNFAIP3	MYO10	BIRC3
BCL10	CARD9	CASP8	BIRC2
MYH2	MYO1C	CARD11	BAX
FADD	MYO1G	BCL2	IL2

Organ development GO0048515

IL18	MYH6	KIF3A	MYO7A
IL10	MYH9	SPHK2	CASP5
BAK1	MYO18B	MYO1E	CASP9
MYO6	RIPK2	MALT1	BOK
MYO3A	BCL2L2	HDAC1	BAG3
MYH3	BCL2L1	MYH11	CASP8
DFFB	MYO15A	MYH10	CASP2
MYH7	SPP1	RTN4	ACTA1
FADD	CFLAR	DEDD	ACTA2
ACTB	DNAH5	MYO5A	CARD11
ACTC	BCL2	BID	FAIM2
BIRC6	BAX	BIK	APAF1
IL2			

FatiGO

Nucleic Acids Research, 2007, Vol. 35, Web Server issue **W91–W96**
doi:10.1093/nar/gkm260

FatiGO+: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments

BIOINFORMATICS **APPLICATIONS NOTE**

FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes

*Fátima Al-Shahrour, Ramón Díaz-Uriarte and Joaquín Dopazo**

Bioinformatics Unit, Centro Nacional de Investigaciones Oncológicas (CNIO), Melchor Fernández Almagro 3, 28029 Madrid, Spain

Received on August 15, 2003; revised on September 30, 2003; accepted on October 1, 2003
Advance Access publication January 22, 2004

ABSTRACT

Summary: We present a simple but powerful procedure to extract Gene Ontology (GO) terms that are significantly over- or under-represented in sets of genes within the context of a genome-scale experiment (DNA microarray, proteomics, etc.). Said procedure has been implemented as a web application, FatiGO, allowing for easy and interactive querying. FatiGO, which takes the multiple-testing nature of statistical contrast into account, currently includes GO associations for diverse organisms (human, mouse, fly, worm and yeast) and the TrEMBL/Swissprot GOAnnotations@EBI correspondences from the European Bioinformatics Institute.

Availability: http://fatigo.bioinfo.cnio.es

Contact: jdopazo@cnio.es

Most resources available that collect information regarding gene or protein function, biological properties, etc., are based on the pre-genomic design in which the information is accessed and displayed in a one-gene-at-a-time format. Nevertheless, many problems related to functional genomics involve the detection of biological properties, functions, etc., shared by a set of genes, that sets them aside from the remaining ones. The practical application of methods of pre-genomic design to these problems is drastically limited when thousands of genes are involved in the comparative study. The use of techniques of automatic management of biological information, such as text mining, in studying the coherence of gene groups obtained from different methodologies has only recently been addressed (Oliveros *et al.*, 2000; Raychaudhuri *et al.*, 2002; Pavlidis *et al.*, 2002), although its practical application still has many drawbacks (Blaschke *et al.*, 2002). Furthermore, real implementations are often scarce and beyond the reach of many users.

An alternative to extracting information from scientific text sources is by using ontologies. In its most simple representation, ontologies provide a structured description of biological information that is extremely useful for computational management. One of the most widely accepted ontologies is Gene Ontology (GO; Ashburner *et al.*, 2000), which organizes information for molecular function, biological processes and cellular components for a number of different organisms. The potential of GO terms as a structured source of information however, has yet to be fully exploited.

Here we present FatiGO, a web-based application (http://fatigo.bioinfo.cnio.es). Since the publication of FatiGO in the GO consortium web page (http://www.geneontology.org) less than a year ago, a number of tools have been implemented based on the same idea of mapping biological knowledge on sets of genes. Thus, Onto-Express (Khatri *et al.*, 2002), which generates tables that correlate groups of genes to biochemical and molecular functions or MAPFinder (Doniger *et al.*, 2003), which, using a searchable web interface, identifies GO terms over-represented in the data. A similar tool, FunSpec (Robinson *et al.*, 2002), evaluates groups of yeast genes in terms of their annotations in diverse databases. Many of these tools are stand-alone applications with user-friendly interfaces, but obviously suppose limitations in processing large amounts of data. Moreover, important issues such as the multiple-testing nature of the statistical contrasts are not well addressed.

FatiGO is used to extract relevant GO terms for a group of genes with respect to a set of genes of reference (typically the rest of genes). The terms are considered to be relevant by the application of a Fisher's exact test that considers the multiple-testing nature of the statistical contrast performed. Multiple testing is an important issue that is, nevertheless, scarcely addressed (Slonim, 2002). If the multiple-testing nature of the statistical contrast is not taken into account an increase in the rate of false positives (i.e. terms identified as over- or under-represented whose proportions, in reality, are not significantly different), occurs. FatiGO can deal with thousands of genes from different organisms (currently human, mouse, *Drosophila*, worm, yeast, as well as genes whose proteins are included in Swissprot database), and can be queried using

*To whom correspondence should be addressed.

578

Annotation	List 1	List 2
GO:0006915	5	76
GO:0002684	6	14
GO:0048515	19	30

FatiGO

Nucleic Acids Research, 2007, Vol. 35, Web Server issue W91–W96
doi:10.1093/nar/gkm260

FatiGO+: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments

BIOINFORMATICS **APPLICATIONS NOTE**

FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes

*Fátima Al-Shahrour, Ramón Díaz-Uriarte and Joaquín Dopazo**

Bioinformatics Unit, Centro Nacional de Investigaciones Oncológicas (CNIO), Melchor Fernández Almagro 3, 28029 Madrid, Spain

Received on August 15, 2003; revised on September 30, 2003; accepted on October 1, 2003
Advance Access publication January 22, 2004

ABSTRACT

Summary: We present a simple but powerful procedure to extract Gene Ontology (GO) terms that are significantly over- or under-represented in sets of genes within the context of a genome-scale experiment (DNA microarray, proteomics, etc.). Said procedure has been implemented as a web application, FatiGO, allowing for easy and interactive querying. FatiGO, which takes the multiple-testing nature of statistical contrast into account, currently includes GO associations for diverse organisms (human, mouse, fly, worm and yeast) and the TrEMBL/Swissprot GOAnnotations@EBI correspondences from the European Bioinformatics Institute.

Availability: http://fatigo.bioinfo.cnio.es

Contact: jdopazo@cnio.es

Most resources available that collect information regarding gene or protein function, biological properties, etc., are based on the pre-genomic design in which the information is accorded and displayed in a one-gene-at-a-time format. Nevertheless, many problems related to functional genomics involve the detection of biological properties, functions, etc., shared by a set of genes, that sets them aside from the remaining ones. The practical application of methods of pre-genomic design to these problems is drastically limited when thousands of genes are involved in the comparative study. The use of techniques of automatic management of biological information, such as text mining, in studying the coherence of gene groups obtained from different methodologies has only recently been addressed (Oliveros *et al.*, 2000; Raychaudhuri *et al.*, 2002; Pavlidis *et al.*, 2002), although its practical application still has many drawbacks (Blaschke *et al.*, 2002). Furthermore, real implementations are often scarce and beyond the reach of many users.

An alternative to extracting information from scientific text sources is by using ontologies. In its most simple representation, ontologies provide a structured description of biological information that is extremely useful for computational management. One of the most widely accepted ontologies is Gene Ontology (GO; Ashburner *et al.*, 2000), which organizes information for molecular function, biological processes and cellular components for a number of different organisms. The potential of GO terms as a structured source of information however, has yet to be fully exploited.

Here we present FatiGO, a web-based application (http://fatigo.bioinfo.cnio.es). Since the publication of FatiGO in the GO consortium web page (http://www.geneontology.org) less than a year ago, a number of tools have been implemented based on the same idea of mapping biological knowledge on sets of genes. Thus, Onto-Express (Khatri *et al.*, 2002), which generates tables that correlate groups of genes to biochemical and molecular functions or MAPFinder (Doniger *et al.*, 2003), which, using a searchable web interface, identifies GO terms over-represented in the data. A similar tool, FunSpec (Robinson *et al.*, 2002), evaluates groups of yeast genes in terms of their annotations in diverse databases. Many of these tools are stand-alone applications with user-friendly interfaces, but obviously suppose limitations in processing large amounts of data. Moreover, important issues such as the multiple-testing nature of the statistical contrasts are not well addressed.

FatiGO is used to extract relevant GO terms for a group of genes with respect to a set of genes of reference (typically the rest of genes). The terms are considered to be relevant by the application of a Fisher's exact test that considers the multiple-testing nature of the statistical contrast performed. Multiple testing is an important issue that is, nevertheless, scarcely addressed (Slonim, 2002). If the multiple-testing nature of the statistical contrast is not taken into account an increase in the rate of false positives (i.e. terms identified as over- or under-represented whose proportions, in reality, are not significantly different), occurs. FatiGO can deal with thousands of genes from different organisms (currently human, mouse, *Drosophila*, worm, yeast, as well as genes whose proteins are included in Swissprot database), and can be queried using

*To whom correspondence should be addressed.

578

Annotation	List 1	List 2
GO:0006915	5	76
GO:0002684	6	14
GO:0048515	19	30

Fisher exact test

Annotation	LOR	*p value*	FDR
GO:0006915	-4.13	$8.41*10^{-30}$	$3.67•10^{-27}$
GO:0002684	-1.11	0.033	0.049
GO:0048515	-0.79	0.016	0.029

FatiGO

Nucleic Acids Research, 2007, Vol. 35, Web Server issue W91–W96
doi:10.1093/nar/gkm260

FatiGO+: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments

BIOINFORMATICS **APPLICATIONS NOTE**

FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes

*Fátima Al-Shahrour, Ramón Díaz-Uriarte and Joaquín Dopazo**

Bioinformatics Unit, Centro Nacional de Investigaciones Oncológicas (CNIO), Melchor Fernández Almagro 3, 28029 Madrid, Spain

Received on August 15, 2003; revised on September 30, 2003; accepted on October 1, 2003
Advance Access publication January 22, 2004

ABSTRACT

Summary: We present a simple but powerful procedure to extract Gene Ontology (GO) terms that are significantly over- or under-represented in sets of genes within the context of a genome-scale experiment (DNA microarray, proteomics, etc.). Said procedure has been implemented as a web application, FatiGO, allowing for easy and interactive querying. FatiGO, which takes the multiple-testing nature of statistical contrast into account, currently includes GO associations for diverse organisms (human, mouse, fly, worm and yeast) and the TrEMBL/Swissprot GOAnnotations@EBI correspondences from the European Bioinformatics Institute.

Availability: http://fatigo.bioinfo.cnio.es

Contact: jdopazo@cnio.es

Most resources available that collect information regarding gene or protein function, biological properties, etc., are based on the pre-genomic design in which the information is accorded and displayed in a one-gene-at-a-time format. Nevertheless, many problems related to functional genomics involve the detection of biological properties, functions, etc., shared by a set of genes, that sets them aside from the remaining ones. The practical application of methods of pre-genomic design to these problems is drastically limited when thousands of genes are involved in the comparative study. The use of techniques of automatic management of biological information, such as text mining, in studying the coherence of gene groups obtained from different methodologies has only recently been addressed (Oliveros *et al.*, 2000; Raychaudhuri *et al.*, 2002; Pavlidis *et al.*, 2002), although its practical application still has many drawbacks (Blaschke *et al.*, 2002). Furthermore, real implementations are often scarce and beyond the reach of many users.

An alternative to extracting information from scientific text sources is by using ontologies. In its most simple representation, ontologies provide a structured description of biological information that is extremely useful for computational management. One of the most widely accepted ontologies is Gene Ontology (GO; Ashburner *et al.*, 2000), which organizes information for molecular function, biological processes and cellular components for a number of different organisms. The potential of GO terms as a structured source of information however, has yet to be fully exploited.

Here we present FatiGO, a web-based application (http://fatigo.bioinfo.cnio.es). Since the publication of FatiGO in the GO consortium web page (http://www.geneontology.org) less than a year ago, a number of tools have been implemented based on the same idea of mapping biological knowledge on sets of genes. Thus, Onto-Express (Khatri *et al.*, 2002), which generates tables that correlate groups of genes to biochemical and molecular functions or MAPFinder (Doniger *et al.*, 2003), which, using a searchable web interface, identifies GO terms over-represented in the data. A similar tool, FunSpec (Robinson *et al.*, 2002), evaluates groups of yeast genes in terms of their annotations in diverse databases. Many of these tools are stand-alone applications with user-friendly interfaces, but obviously suppose limitations in processing large amounts of data. Moreover, important issues such as the multiple-testing nature of the statistical contrasts are not well addressed.

FatiGO is used to extract relevant GO terms for a group of genes with respect to a set of genes of reference (typically the rest of genes). The terms are considered to be relevant by the application of a Fisher's exact test that considers the multiple-testing nature of the statistical contrast performed. Multiple testing is an important issue that is, nevertheless, scarcely addressed (Slonim, 2002). If the multiple-testing nature of the statistical contrast is not taken into account an increase in the rate of false positives (i.e. terms identified as over- or under-represented whose proportions, in reality, are not significantly different), occurs. FatiGO can deal with thousands of genes from different organisms (currently human, mouse, *Drosophila*, worm, yeast, as well as genes whose proteins are included in Swissprot database), and can be queried using

*To whom correspondence should be addressed.

578

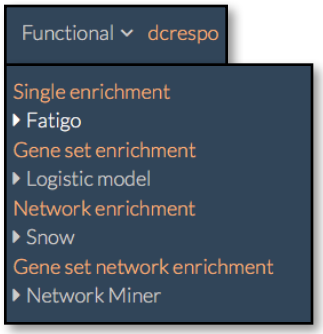
Annotation	List 1	List 2
GO:0006915	5	76
GO:0002684	6	14
GO:0048515	19	30

Fisher exact test

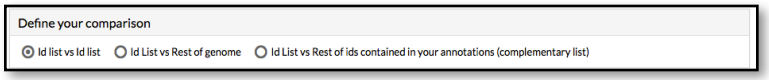
Annotation	LOR	*p value*	FDR
GO:0006915	-4.13	$8.41*10^{-30}$	$3.67•10^{-27}$
GO:0002684	-1.11	0.033	0.049
GO:0048515	-0.79	0.016	0.029

FatiGO on *Babelomics 5*

Tool selection



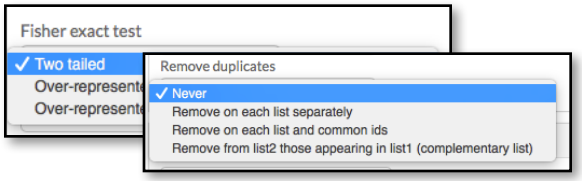
Comparison



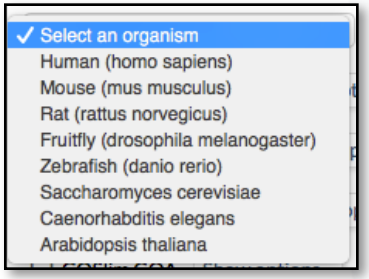
First list selection

Second list selection

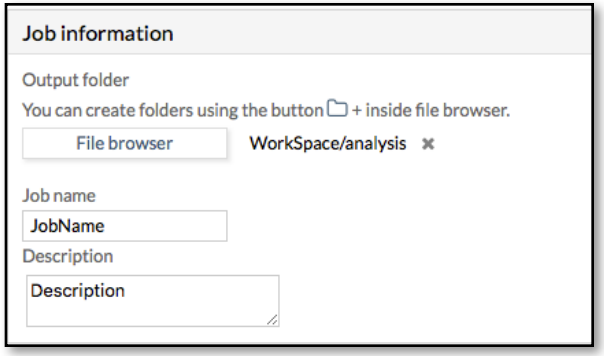
Test options

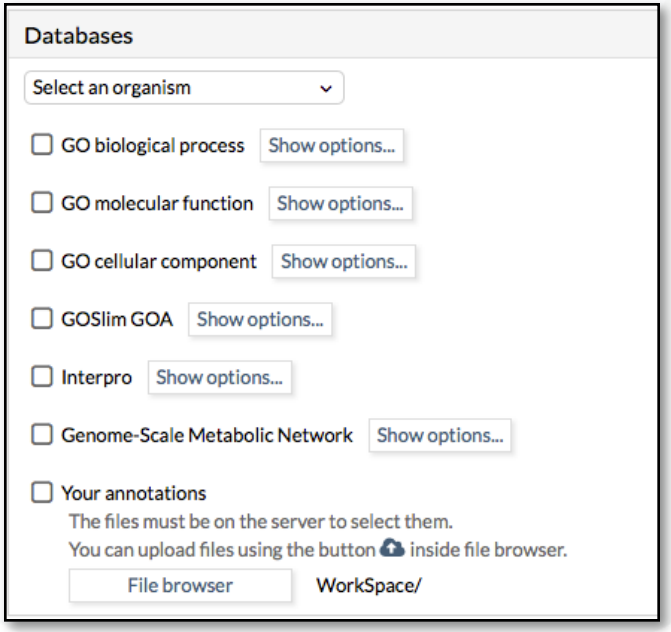


Databases



Job information





Data input

Gene list: bioformat"ID"

```
MBP-de-Dan-2:example datocresbe$
MYO5B
DNAL4
MYO18B
MYH9
C20orf23
RPLP0P1
RPLP0P1
DNCL2A
KIF3B
MYH7B
MYH6
MYH6
MYH7
KIF13B
KIF26A
KNS2
DNCH1
MYO9A
MYO5C
MYO5A
MYO1E
```

1 Column: Gene ID

Custom annotation: bioformat Gene vs annotation

```
MBP-de-Dan-2:example datocresbe$ head -30 ../annotation
"rno04014__47 49" "plasma membrane"
"rno04014__47 49" "positive regulation of cell prolifer
"rno04014__47 49" "extrinsic component of cytoplasmic s
"rno04014__47 49" "GTP binding"
"rno04014__47 49" "Golgi apparatus"
"rno04014__47 49" "positive regulation of angiogenesis"
"rno04014__47 49" "molecular_function"
"rno04014__47 49" "biological_process"
"rno04014__47 49" "cellular_component"
"rno04014__42" "lipid catabolic process"
"rno04014__42" "phospholipase A2 activity"
"rno04014__42" "molecular_function"
"rno04014__42" "mitochondrial inner membrane"
"rno04014__42" "cellular response to insulin stimulus"
"rno04014__42" "glucose transport"
"rno04014__43" "RNA polymerase II transcription factor
"rno04014__43" "transcription from RNA polymerase II pr
"rno04014__43" "cell differentiation"
"rno04014__44" "nucleus"
"rno04014__44" "RNA polymerase II transcription factor
"rno04014__44" "transcription from RNA polymerase II pr
```

Column 1: Gene ID

Column 2: Function

Reading results

Analysis overview

Job information

Name : *motor vs apop*
Description : *Description*
Tool : *fatigo*
Output folder : *WorkSpace/analysis/20160224155426/*

Input data

Species: *hsa*
Duplicates management: *Never remove*
Fisher exact test: *Two tailed*
List 1 (after duplicates managing) *clean_list1.txt*
List 2 (after duplicates managing) *clean_list2.txt*

Summary

Id annotations per DB *annotations_per_db.txt*

#DB	List1	List 2
GO biological process propagated	85 of 124 (68.55%) 63.66 annotations/id	79 of 105 (75.24%) 99.28 annotations/id

1 Results ‹ 1 of 1 ›

Results

Significant Results

Number of significant terms per DB *significant_count_0.05.txt*

Select p-value: 0.1 | 0.05 | 0.01 | 0.005

#DB	N° of significant terms
GO biological process propagated	588

1 Results ‹ 1 of 1 ›

GO biological process propagated

GO biological process propagated significant terms (pvalue<0.05) *significant_go_biological_process_propagated_0.05.txt*

Term	Term size	Term size(in genome)	Annotations lists	Annotated ids list	Odds ratio (log e)	Pvalue	Adj. Pvalue
protein secretion(GO:0009306)	5	500	List 1: 0% List 2: 4.76%	List 1: List 2: IL10 CIDEA CASP5 CASP1	-1.7976931349e+304	0.019	0.03
regulation of release of cytochrome c from mitochondria(GO:0090199)	8	139	List 1: 0.81% List 2: 6.67%	List 1: OPA1 List 2: BAK1 HRK BCL2L1 BNIP3	-2.17	0.025	0.039
positive regulation of protein ubiquitination(GO:0031398)	6	376	List 1: 0% List 2: 5.71%	List 1: List 2: BCL10 RIPK2 MALT1 BRCA1	-1.7976931349e+304	0.0086	0.017
regulation of protein modification process(GO:0031399)	21	2965	List 1: 1.61% List 2: 18.1%	List 1: DCTN1 CENPE List 2: BAK1 BCL10 RIPK2 TNFAIP3	-2.6	1.2e-5	6.7e-5
regulation of protein ubiquitination(GO:0031396)	8	559	List 1: 0% List 2: 7.62%	List 1: List 2: BCL10 RIPK2 TNFAIP3 MALT1	-1.7976931349e+304	0.0017	0.0046
immune system development(GO:0002520)	14	1449	List 1: 1.61% List 2: 11.43%	List 1: MYH9 MYO1E List 2: IL10 BAK1 FADD RIPK2	-2.06	0.0039	0.0089
leukocyte differentiation(GO:0002521)	11	815	List 1: 0.81% List 2: 9.52%	List 1: MYH9 List 2: IL10 BAK1 FADD RIPK2	-2.56	0.0031	0.0081
positive regulation of intracellular signal transduction(GO:1902533)	22	1709	List 1: 0% List 2: 20.95%	List 1: List 2: BCL10 FADD RIPK2	-1.7976931349e+304	9.4e-9	8.9e-8

Download results

All results

GO biological process propagated *go_biological_process_propagated.txt*

Annotation files

Annotations for GO biological process propagated *go_biological_process_propagated.annot*

Reading results

Analysis overview

Job information

Name : *motor vs apop*
Description : *Description*
Tool : *fatigo*
Output folder : *WorkSpace/analysis/20160224155426/*

Summary

Id annotations per DB *annotations_per_db.txt*

#DB	List1	List 2
GO biological process propagated	85 of 124 (68.55%) 63.66 annotations/id	79 of 105 (75.24%) 99.28 annotations/id

1 Results ‹ 1 of 1 ›

Input data

Species: *hsa*
Duplicates management: *Never remove*
Fisher exact test: *Two tailed*
List 1 (after duplicates managing) *clean_list1.txt*
List 2 (after duplicates managing) *clean_list2.txt*

Results

Significant Results

Number of significant terms per DB *significant_count_0.05.txt*

Select p-value: 0.1 | 0.05 | 0.01 | 0.005

#DB	N° of significant terms
GO biological process propagated	588

1 Results ‹ 1 of 1 ›

GO biological process propagated

GO biological process propagated significant terms (pvalue<0.05) *significant_go_biological_process_propagated_0.05.txt*

Term	Term size	Term size(in genome)	Annotations lists	Annotated ids list		Odds ratio (log e)	Pvalue	Adj. Pvalue
protein secretion(GO:0009306)	5	500	List 1: 0% List 2: 4.76%	List 1:	List 2: IL10 CIDEA CASP5 CASP1	-1.7976931349e+304	0.019	0.03
regulation of release of cytochrome c from mitochondria(GO:0090199)	8	139	List 1: 0.81% List 2: 6.67%	List 1: OPA1	List 2: BAK1 HRK BCL2L1 BNIP3	-2.17	0.025	0.039
positive regulation of protein ubiquitination(GO:0031398)	6	376	List 1: 0% List 2: 5.7%	List 1:	List 2: BCL10 RIPK2 MALT1 BRCA1	-1.7976931349e+304	0.0086	0.017
regulation of protein modification process(GO:0031399)	21	2965	List 1: 1.61% List 2: 18.1%	List 1: DCTN1 CENPE	List 2: BAK1 BCL10 RIPK2 TNFAIP3	-2.6	1.2e-5	6.7e-5
regulation of protein ubiquitination(GO:0031396)	8	559	List 1: 0% List 2: 7.62%	List 1:	List 2: BCL10 RIPK2 TNFAIP3 MALT1	-1.7976931349e+304	0.0017	0.0046
immune system development(GO:0002520)	14	1449	List 1: 1.61% List 2: 11.43%	List 1: MYH9 MYO1E	List 2: IL10 BAK1 FADD RIPK2	-2.06	0.0039	0.0089
leukocyte differentiation(GO:0002521)	11	815	List 1: 0.81% List 2: 9.52%	List 1: MYH9	List 2: IL10 BAK1 FADD RIPK2	-2.56	0.0031	0.0081
positive regulation of intracellular signal transduction(GO:1902533)	22	1709	List 1: 0% List 2: 20.95%	List 1:	List 2: BCL10 FADD RIPK2	-1.7976931349e+304	9.4e-9	8.9e-8

Enriched class

Annotated genes from each list

False Discovery Rate

Download results

All results

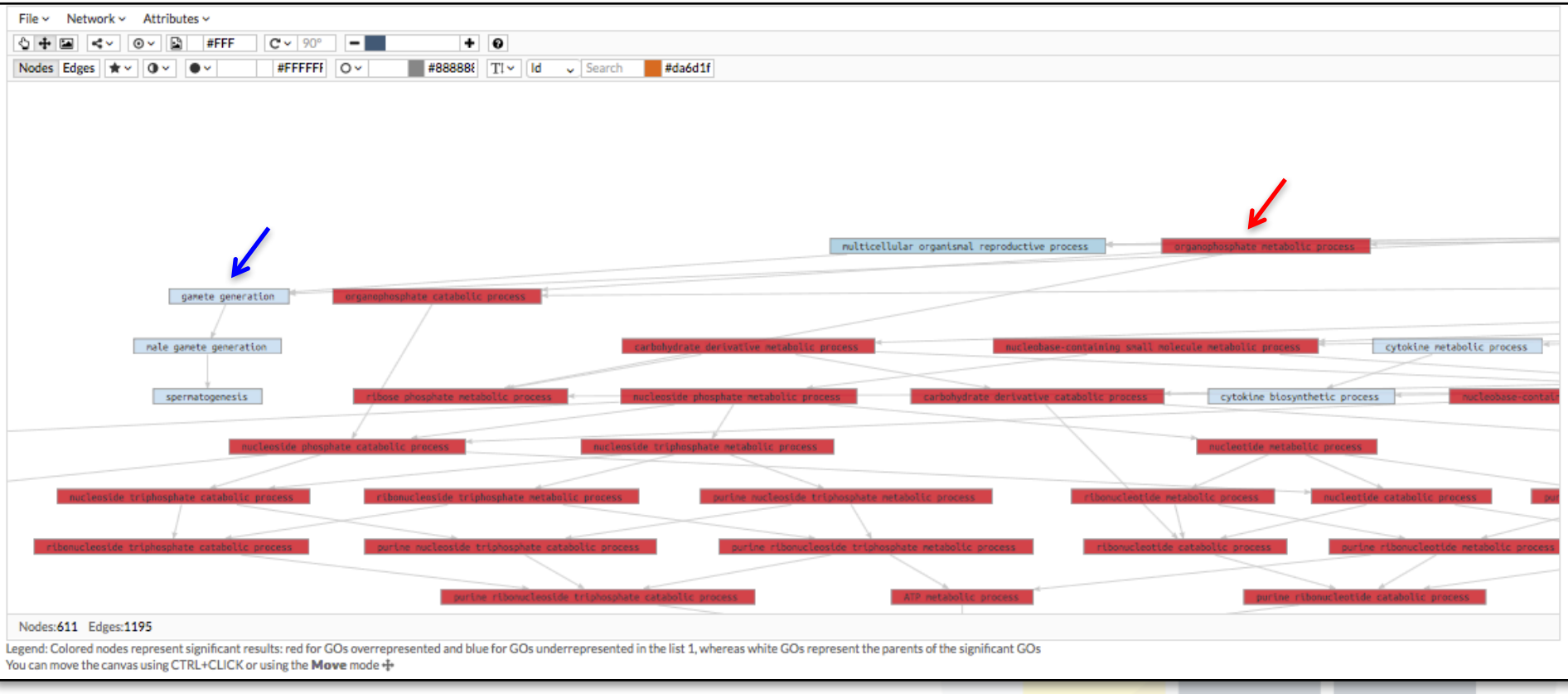
GO biological process propagated *go_biological_process_propagated.txt*

Annotation files

Annotations for GO biological process propagated *go_biological_process_propagated.annot*

Reading results

Results visualization



Legend: Colored nodes represent significant results: red for GOs overrepresented and blue for GOs underrepresented in the list 1, whereas white GOs represent the parents of the significant GOs
You can move the canvas using CTRL+CLICK or using the **Move** mode ✣

Babelomics 5: Gene Set Analysis

GDA

International Course on
Genomic Data Analysis

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

Gene set enrichment

Comparison between a **ranked list** with user defined annotations

Gene set enrichment

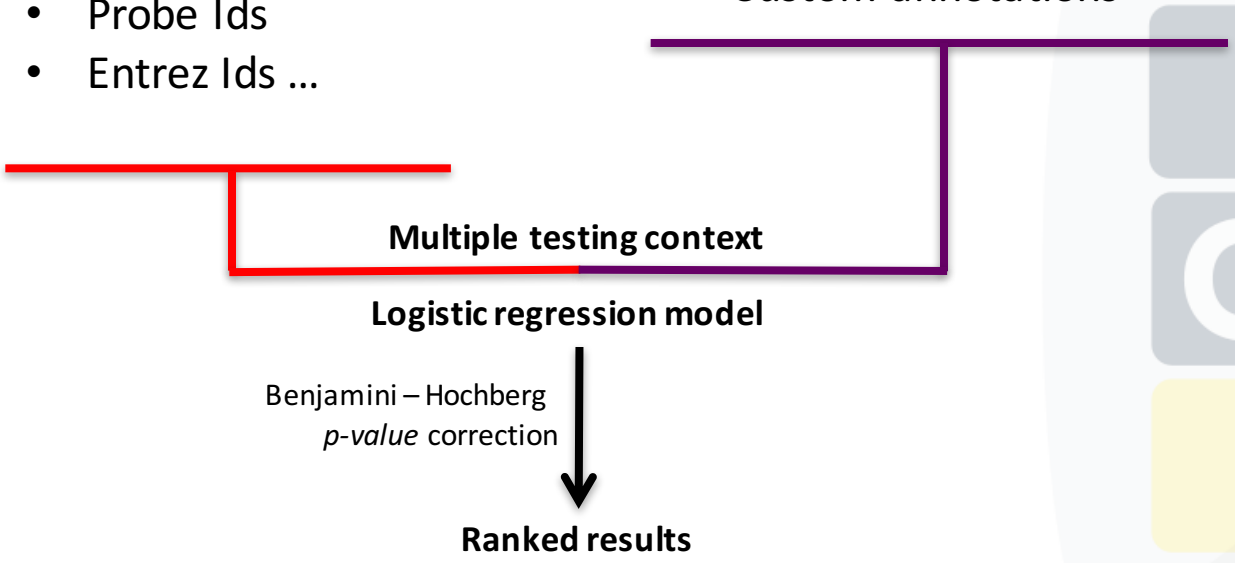
Comparison between a **ranked list** with user defined annotations

Ranked List: User provided

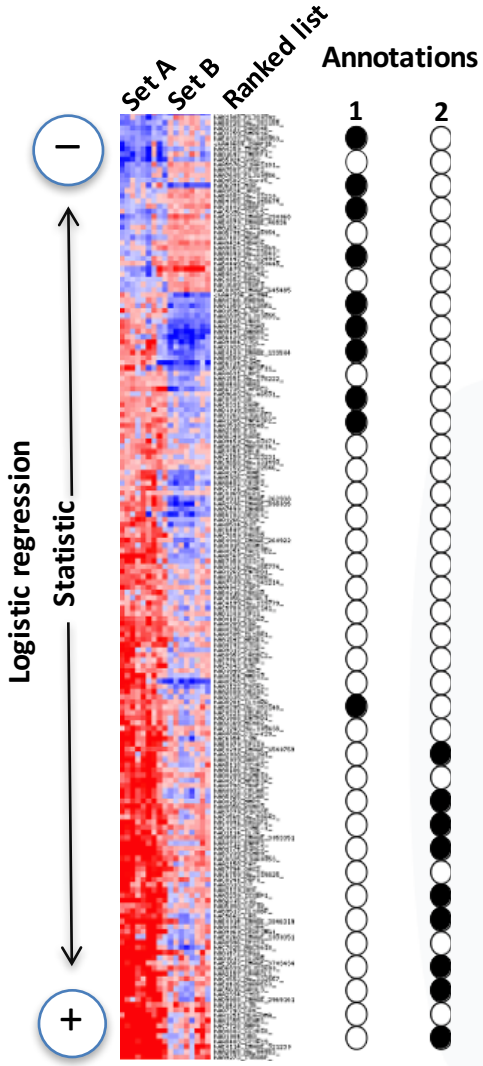
- Gene symbols
- Probe Ids
- Entrez Ids …

Functional annotations:

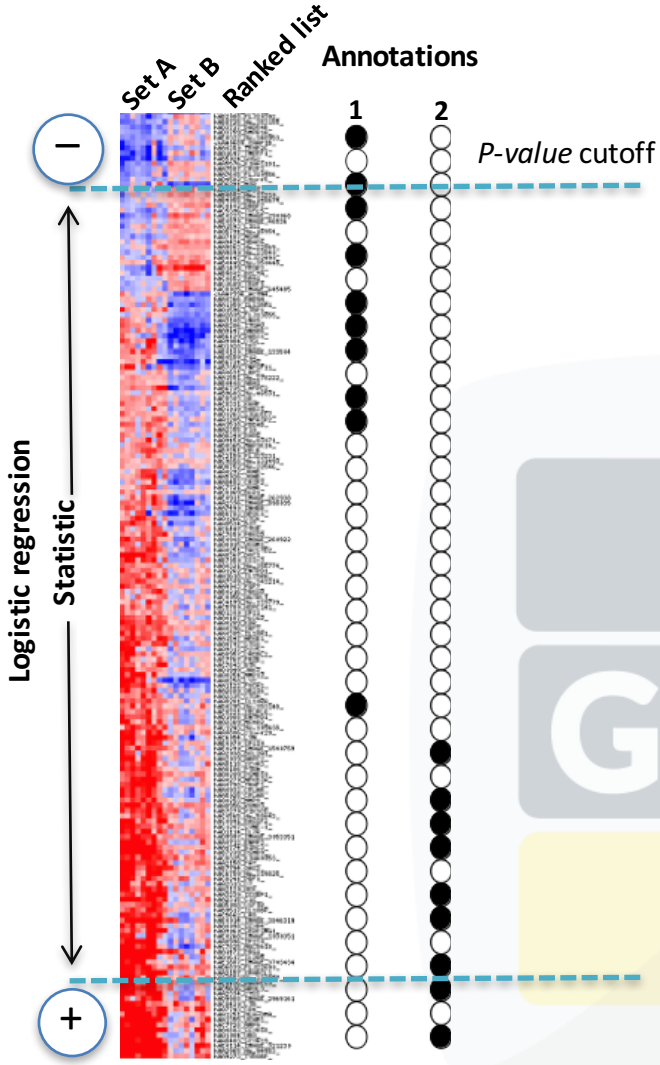
- Pre-defined annotations
- Custom annotations



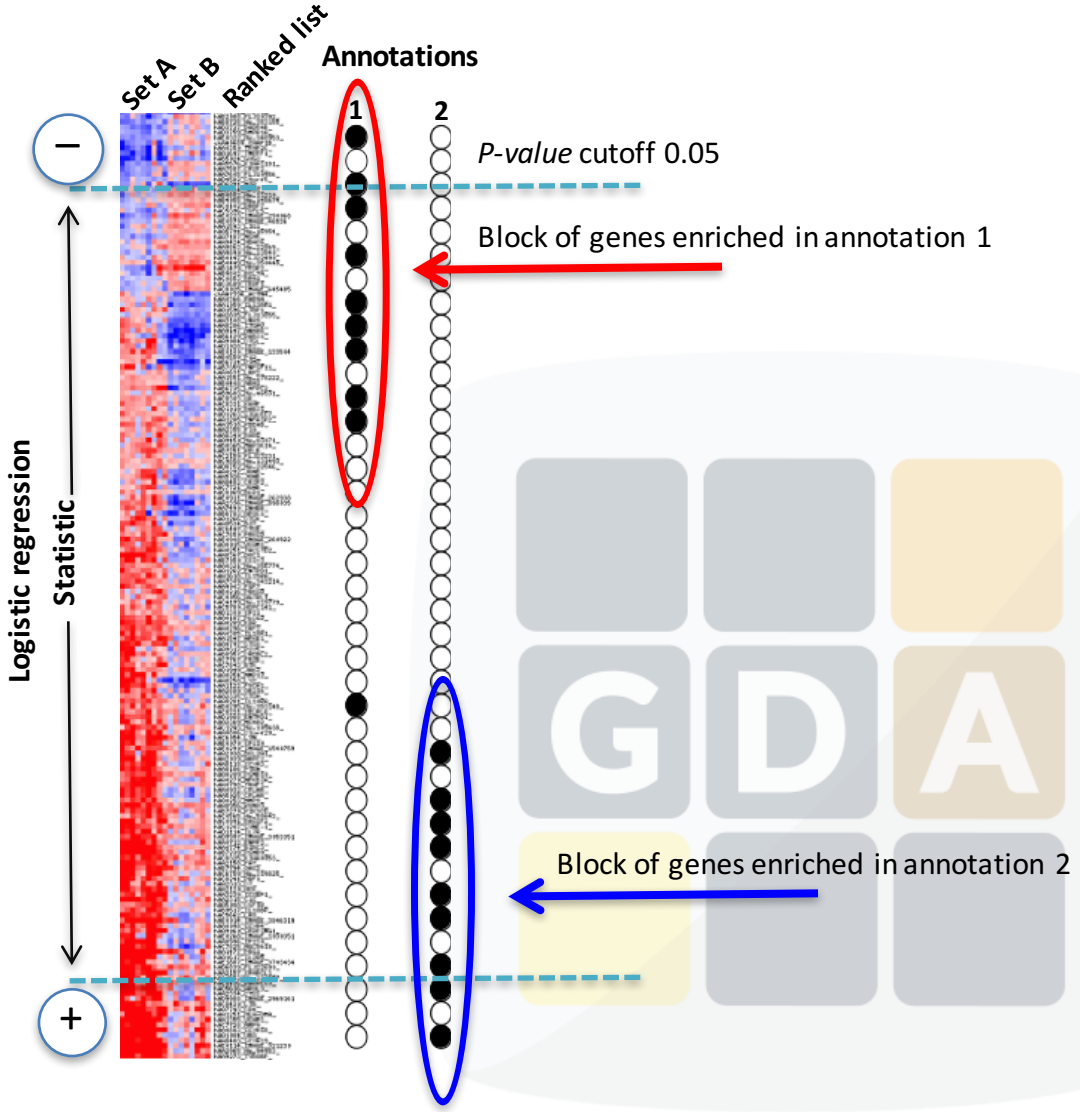
GSA: Logistic model



GSA: Logistic model

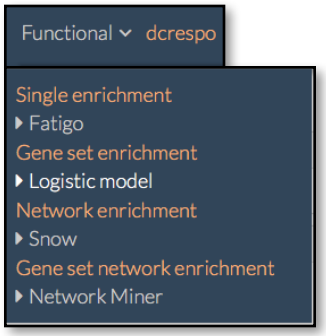


GSA: Logistic model

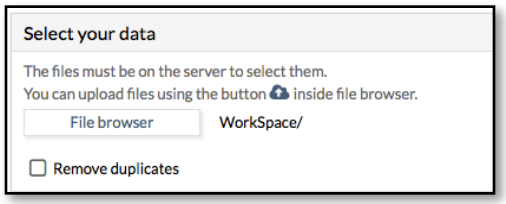


GSA on *Babelomics 5*

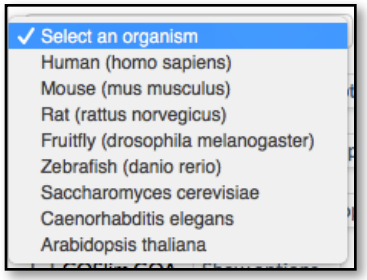
Tool selection

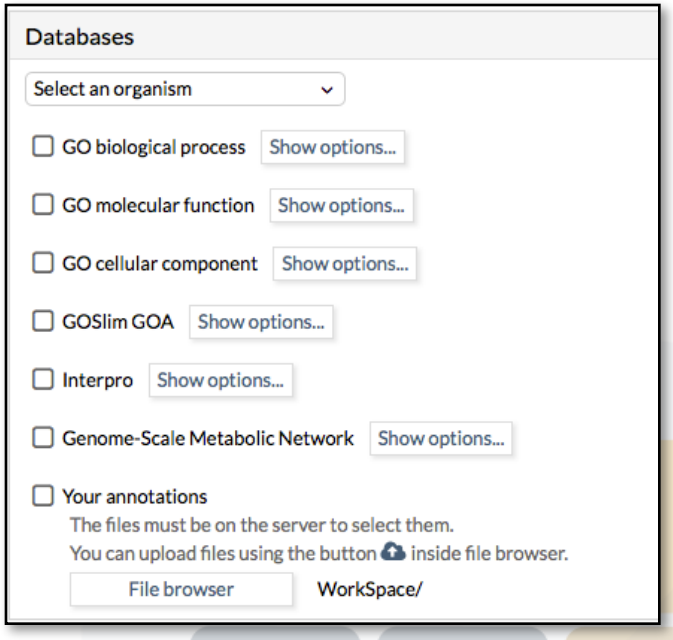


Data selection

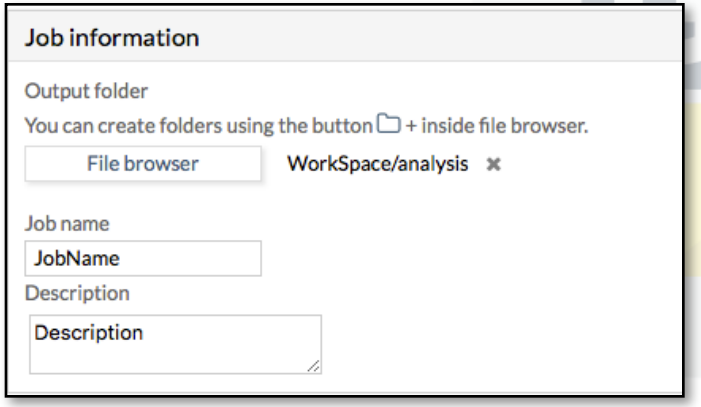


Databases





Job information



Data input

Gene list: bioformat"ID"

```
[MBP-de-Dan-2:data_dan
DYNLT1   14.29
RAB13    9.42
COMMD6   7.3
RPL35    6.99
ZNF124   6.74
NDUFA4   6.74
COX7A2   6.67
UBR2     6.66
YWHAQ    6.65
JAK1     6.65
USMG5    6.51
NDUFA12  6.44
AKT3     6.43
SARS     6.31
STRN     6.28
STK4     6.24
COX6A1   6.14
PIK3R4   6.08
NDUFS4   6.08
FBXO38   6.05
VPS8     6.01
CASP4    5.99
SPATS2L  5.97
IL6ST    5.93
```

Column 1: Ranked Gene ID

Column 2: Rank

Custom annotation: bioformat Gene vs annotation

```
MBP-de-Dan-2:example datocresbe$ head -30 ../annotation
"rno04014__47 49"  "plasma membrane"
"rno04014__47 49"  "positive regulation of cell prolifer
"rno04014__47 49"  "extrinsic component of cytoplasmic s
"rno04014__47 49"  "GTP binding"
"rno04014__47 49"  "Golgi apparatus"
"rno04014__47 49"  "positive regulation of angiogenesis"
"rno04014__47 49"  "molecular_function"
"rno04014__47 49"  "biological_process"
"rno04014__47 49"  "cellular_component"
"rno04014__42"  "lipid catabolic process"
"rno04014__42"  "phospholipase A2 activity"
"rno04014__42"  "molecular_function"
"rno04014__42"  "mitochondrial inner membrane"
"rno04014__42"  "cellular response to insulin stimulus"
"rno04014__42"  "glucose transport"
"rno04014__43"  "RNA polymerase II transcription factor
"rno04014__43"  "transcription from RNA polymerase II pr
"rno04014__43"  "cell differentiation"
"rno04014__44"  "nucleus"
"rno04014__44"  "RNA polymerase II transcription factor
"rno04014__44"  "transcription from RNA polymerase II pr
```

Column 1: Gene ID

Column 2: Function

Reading results

Analysis overview

Job information

Name : *fluorouridine_dataset*
Description : *fluorouridine_dataset*
Tool : *fatiscan*
Output folder : *WorkSpace/analysis/20151120155138/*

Input data

Ranked list *ranked_list.txt*
Id list (sorted) *id_list.txt*
Statistic (sorted) *statistic.txt*

Summary

Id annotations per DB *annotations_per_db.txt*

#DB	Number of annotations
GO biological process propagated	706 of 1000 (70.6%) 59.11 annotations/id

1 Results ‹ 1 of 1 ›

Results

Significant Results

Number of significant terms per DB *significant_count_0.05.txt*

Select p-value: 0.1 | 0.05 | 0.01 | 0.005

#DB	N° of significant terms
GO biological process propagated	209

1 Results ‹ 1 of 1 ›

GO biological process propagated

GO biological process propagated significant terms (pvalue<0.05) *significant_go_biological_process_propagated_0.05.txt*

Term	Term size	Term size(in genome)	annotated_genes lists	converged ids list	lor	adj_pvalue
regulation of myeloid cell differentiation(GO:0045637)	15	355	OGT CSF1 PTK2B ATP6AP1 MEIS2	true	0.28	3.2e-6
mitotic cell cycle(GO:0000278)	51	2418	CCNA1 RCC1 GORASP1 NUDC GAS1	true	0.17	7.3e-6
negative regulation of myeloid cell differentiation(GO:0045638)	7	148	PTK2B MEIS2 INPP4B INPP5D TLR4	true	0.36	1.8e-5
mitotic cell cycle process(GO:1903047)	40	2106	CCNA1 RCC1 NUDC RPS6 NUP205	true	0.17	3.6e-5
single-organism organelle organization(GO:1902589)	91	4963	DLC1 OGT CCNA1 CRYAB RCC1	true	0.13	4.2e-5
regulation of leukocyte differentiation(GO:1902105)	21	435	OGT IL4R IL12A CSF1 ATP6AP1	true	0.21	9.9e-5
RNA metabolic process(GO:0016070)	197	13296	RPL19 HNRPF HOXD11 AGTR2 RPL11	true	0.095	1.5e-4
negative regulation of leukocyte differentiation(GO:1902106)	9	148	IL4R INPP4B INPP5D	true	0.28	1.9e-4

Download results

All results

GO biological process propagated *go_biological_process_propagated.txt*

Annotation files

Annotations for GO biological process propagated *go_biological_process_propagated.annot*

Reading results

Analysis overview

Job information

Name : *fluorouridine_dataset*
Description : *fluorouridine_dataset*
Tool : *fatiscan*
Output folder : *WorkSpace/analysis/20151120155138/*

Input data

Ranked list *ranked_list.txt*
Id list (sorted) *id_list.txt*
Statistic (sorted) *statistic.txt*

Summary

Id annotations per DB *annotations_per_db.txt*

#DB	Number of annotations
GO biological process propagated	706 of 1000 (70.6%) 59.11 annotations/id

1 Results ‹ 1 of 1 ›

Results

Significant Results

Number of significant terms per DB *significant_count_0.05.txt*

Select p-value: 0.1 | 0.05 | 0.01 | 0.005

#DB	N° of significant terms
GO biological process propagated	209

1 Results ‹ 1 of 1 ›

GO biological process propagated

GO biological process propagated significant terms (pvalue<0.05) *significant_go_biological_process_propagated_0.05.txt*

Term	Term size	Term size(in genome)	annotated_genes lists	converged ids list	lor	adj_pvalue
regulation of myeloid cell differentiation(GO:0045637)	15	355	OGT CSF1 PTK2B ATP6AP1 MEIS2	true	0.28	3.2e-6
mitotic cell cycle(GO:0000278)	51	2418	CCNA1 RCC1 GORASP1 NUDC GAS1	true	0.17	7.3e-6
negative regulation of myeloid cell differentiation(GO:0045638)	7	148	PTK2B MEIS2 INPP4B INPP5D TLR4	true	0.36	1.8e-5
[illegible]	49	2106	CCNA1 RCC1 NUDC RPS6 NUP205	true	0.17	3.6e-5
single-organism organelle organization(GO:1902589)	91	4963	DLC1 OGT CCNA1 CRYAB RCC1	true	0.13	4.2e-5
regulation of leukocyte differentiation(GO:1902105)	21	435	OGT IL4R IL12A CSF1 ATP6AP1	true	0.21	9.9e-5
RNA metabolic process(GO:0016070)	197	13296	RPL19 HNRPF HOXD11 AGTR2 RPL11	true	0.095	1.5e-4
negative regulation of leukocyte differentiation(GO:1902106)	9	148	IL4R INPP4B INPP5D [illegible]	true	0.28	1.9e-4

Enriched class

Annotated genes list

False Discovery Rate

Download results

All results

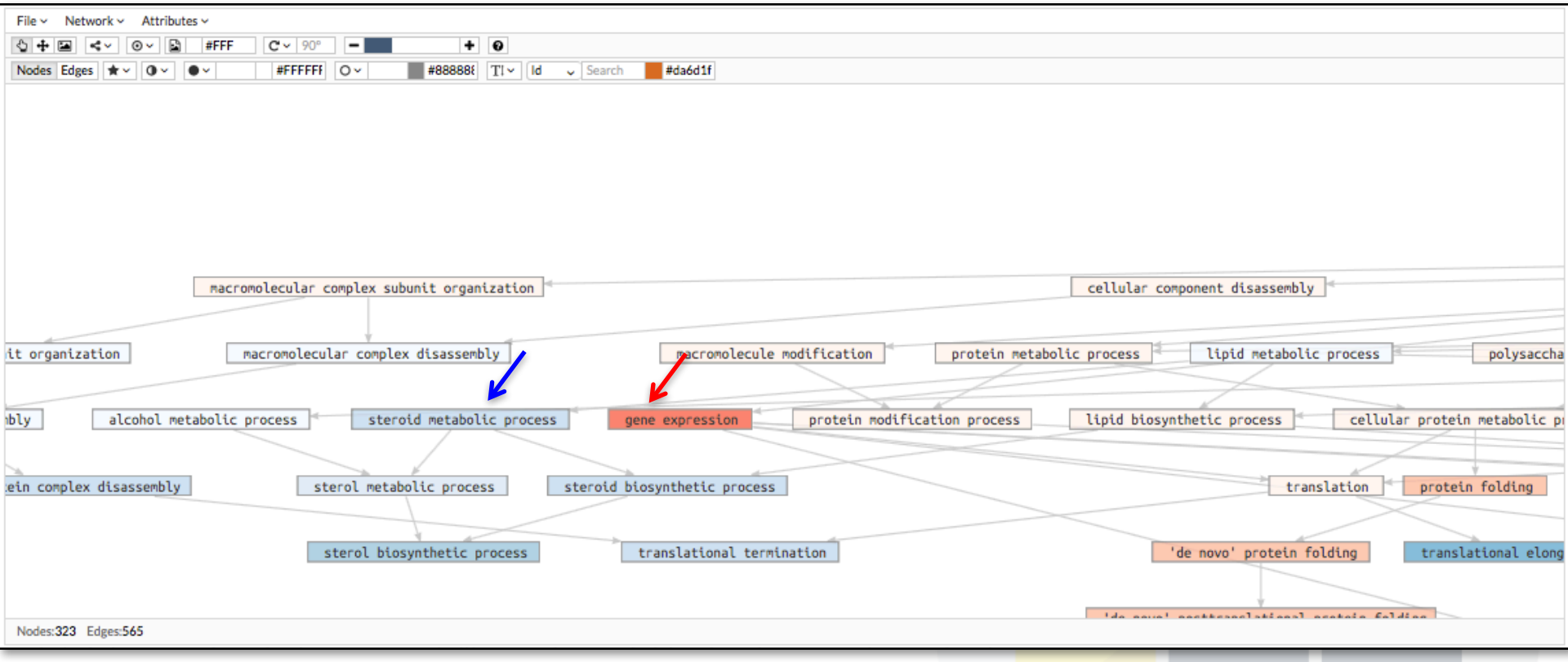
GO biological process propagated *go_biological_process_propagated.txt*

Annotation files

Annotations for GO biological process propagated *go_biological_process_propagated.annot*

Reading results

Results visualization



Keypoints

- **FatiGO:** Detects significant over representation of functional annotations in one gene set respect to the other one.

Keypoints

- **FatiGO:** Detects significant over representation of functional annotations in one gene set respect to the other one.
  - Critical step: Gene list selection. Tipically from differential gene expression.

Keypoints

- **FatiGO:** Detects significant over representation of functional annotations in one gene set respect to the other one.
  - Critical step: Gene list selection. Tipically from differential gene expression.
- **GSA:** Detects gene sets (functional annotations) that are consistently associated to high or low values in a ranked list of genes.

Training data and resources

http://babelomics.org

Any questions?

User:	gda16	gda16b	gda16c
Password:	gda16	gda16b	gda16c

FatiGO

http://www.ncbi.nlm.nih.gov/pubmed/17478504
http://www.ncbi.nlm.nih.gov/pubmed/14990455
wiki: https://github.com/babelomics/babelomics/wiki/Single%20Enrichment

Gene set analysis: Logistic model

http://www.ncbi.nlm.nih.gov/pubmed/20436964
http://www.ncbi.nlm.nih.gov/pubmed/17407596
http://www.ncbi.nlm.nih.gov/pubmed/15840702
wiki: https://github.com/babelomics/babelomics/wiki/Gene%20Set%20Enrichment

Thank you very much for your attention!

Systems genomics team

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

GDA 2016