

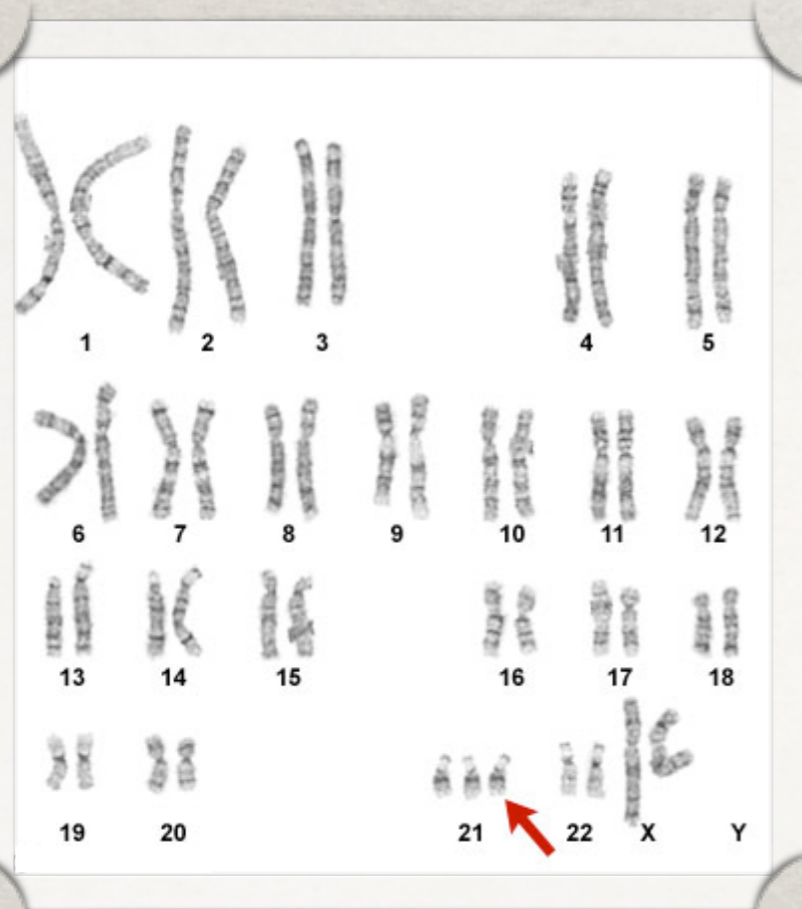
GENE PRIORITIZATION THROUGH GENOMIC DATA FUSION

Léon-Charles Tranchevent, PhD defense, 09/05/2011

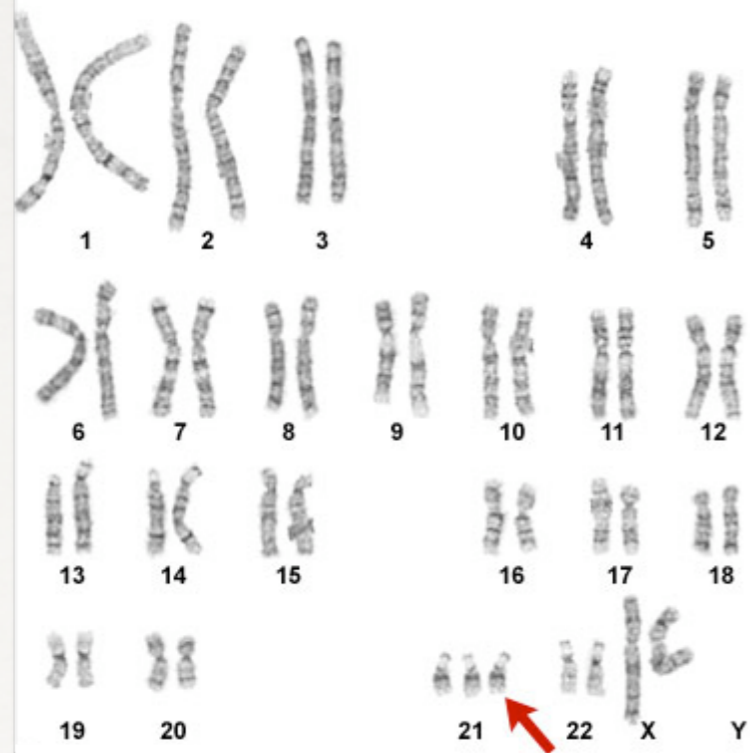
MOTIVATION



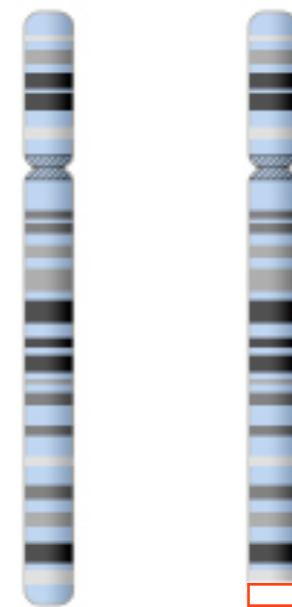
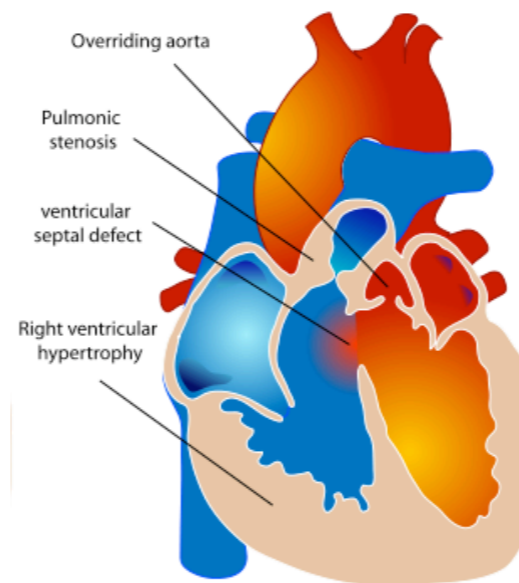
BIOLOGICAL QUESTION



BIOLOGICAL QUESTION

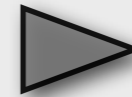
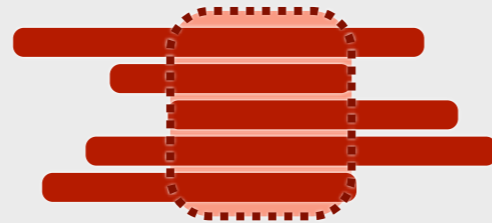
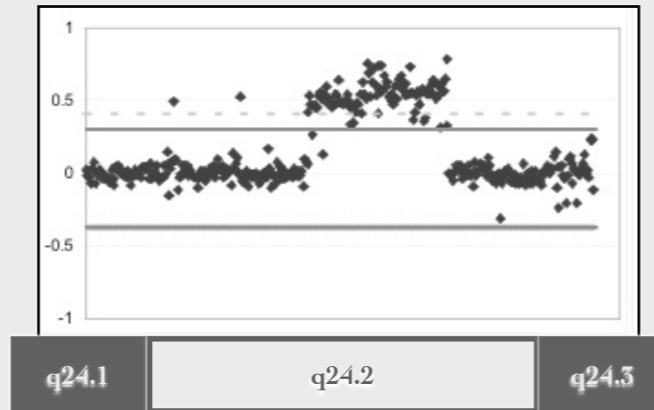
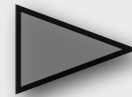
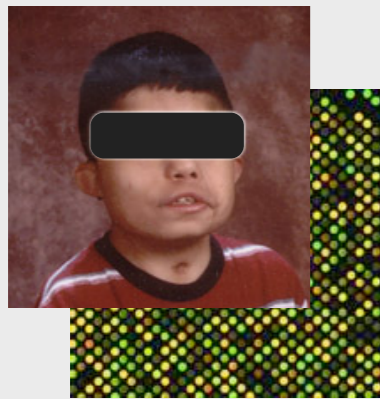


Tetralogy of Fallot



5q deletion

BIOLOGICAL QUESTION

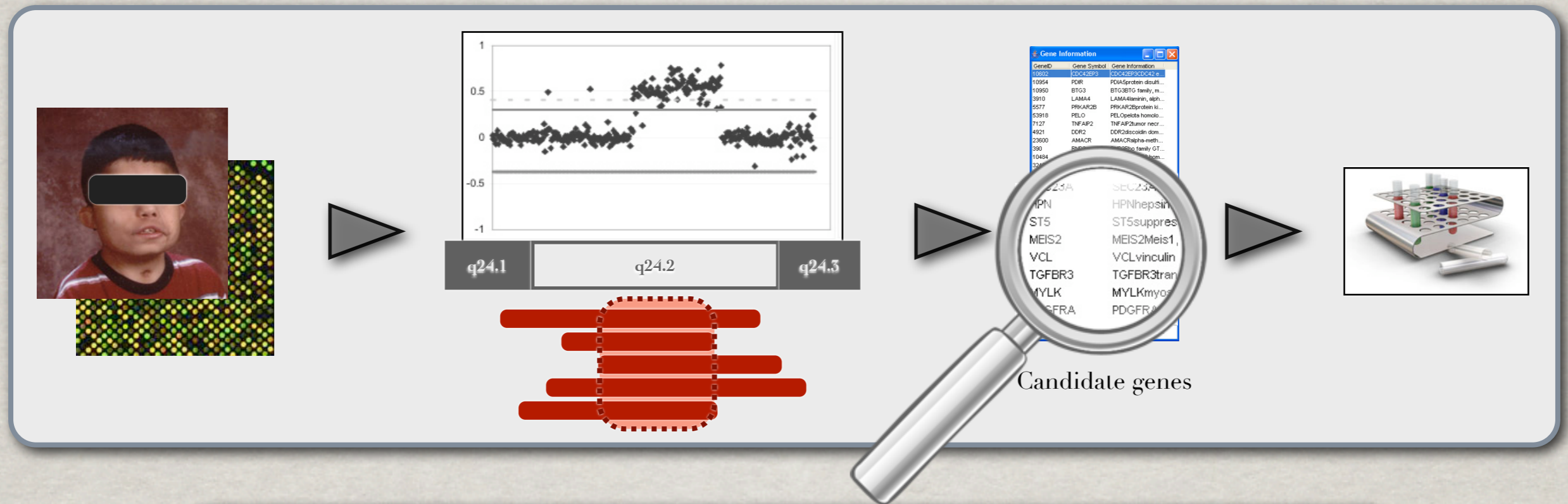


GeneID	Gene Symbol	Gene Information
10602	CDK42EP3	CDK42EP3/CDK42. b...
10954	PDR	PDR/Sproteindisulf...
10950	BTG3	BTG3/STG family, m...
3910	LAMA4	LAMA4/Alonin, alp...
5577	PRKAR2B	PRKAR2B/protein ki...
53918	PELO	PELO/Opeleta homolo...
7127	TNFAP2	TNFAP2/tumor necr...
4921	DDR2	DDR2/discoidin dom...
23600	AMACR	AMACR/Raphe-meth...
390	RND3	RND3/Rho family G1...
10484	SEC23A	SEC23A/Sec23 hom...
3249	HPN	HPN/hepin (transm...
6764	STS	STS/suppression of...
4212	MEIS2	MEIS2/Meis1, myelo...
7414	VCL	VCL/vinculin
7049	TGFB3	TGFB3/transform...
4638	MYLK	MYLK/myosin, light...
5156	PDGFRA	PDGFRA/platelet-de...
57480	PLEKHG1	PLEKHG1/pleckstrin...
10231	DSRH1	DSRH1/Down sy...
6907	TBL1Y	TBL1Y/tetraducin (...)
201254	STRA13	STRA13/stimulated...
2200	FBN1	FBN1/fibrin 1 (Mar...
10020	GNF	GNF/glucosamine (...)
22795	ND2	ND2/ndogen 2 (cost...
6406	SRPX	SRPX/sushi-repeat...
2982	GLUCY1A3	GLUCY1A3/guanylat...
10278	EFS	EFS/embryonal Fyn...
730	C7	C7/complement com...
11215	AKAP11	AKAP11A/kinase (...)
4659	PPP1R12A	PPP1R12A/protein p...
6563	SLC14A1	SLC14A1/solute car...

Candidate genes



BIOLOGICAL QUESTION



Which genes are the most promising candidates?

CONCEPT

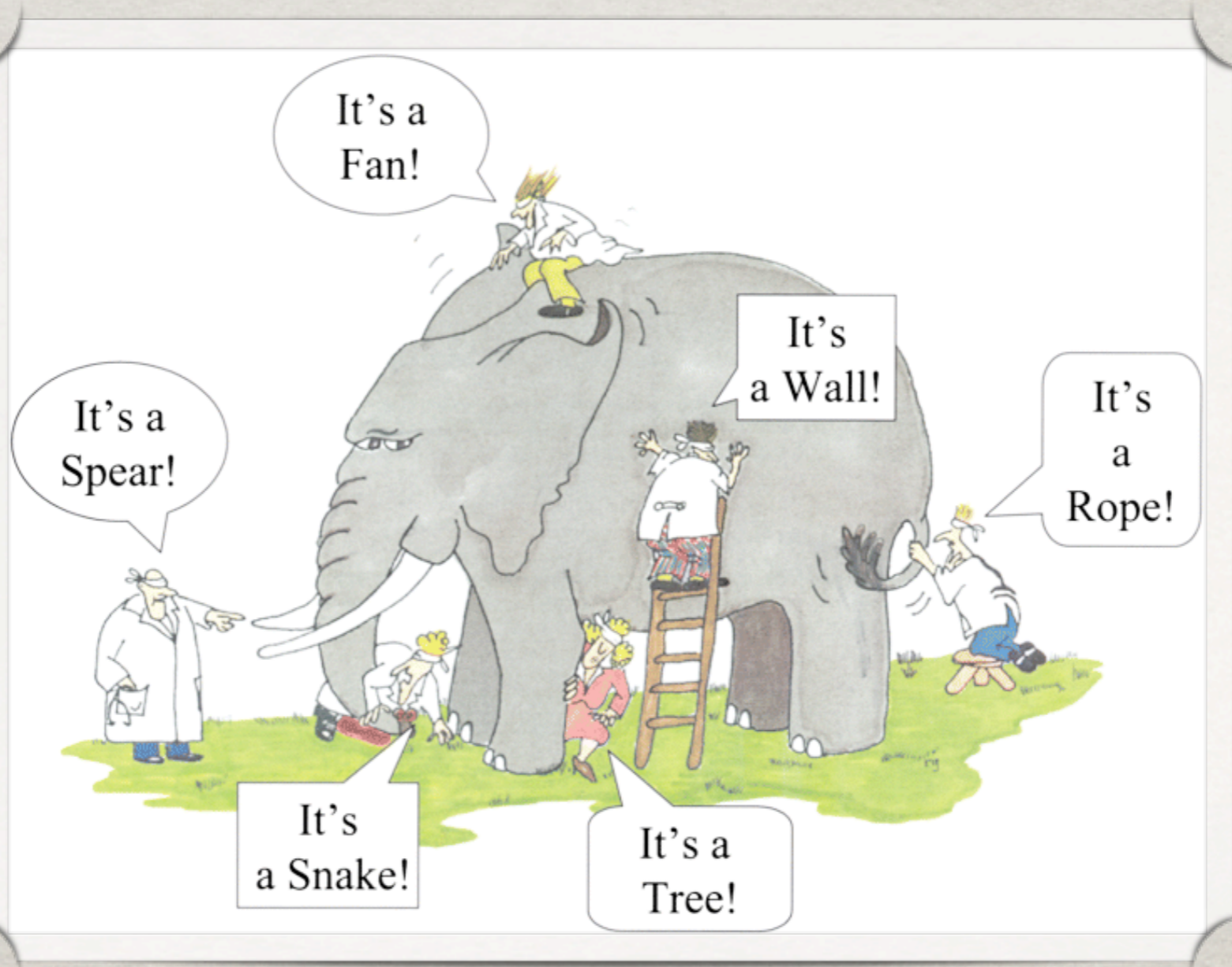
Which genes are the most promising candidates?



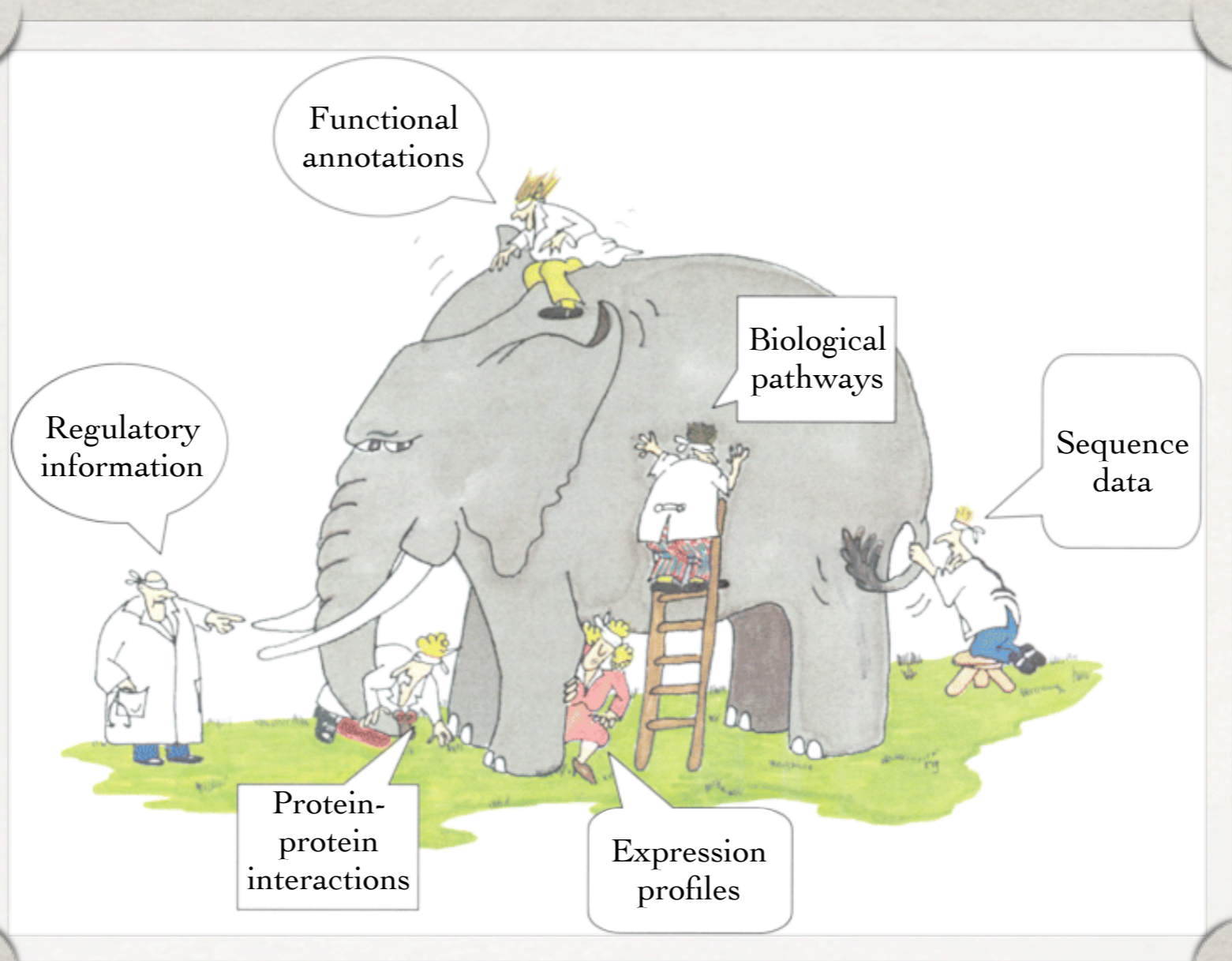
Goh *et al.*, PNAS (2007)
Smith *et al.*, Gene (2003)
Jimenez-Sanchez *et al.*, Nature (2001)

Which genes are similar to the genes involved in the process under study?

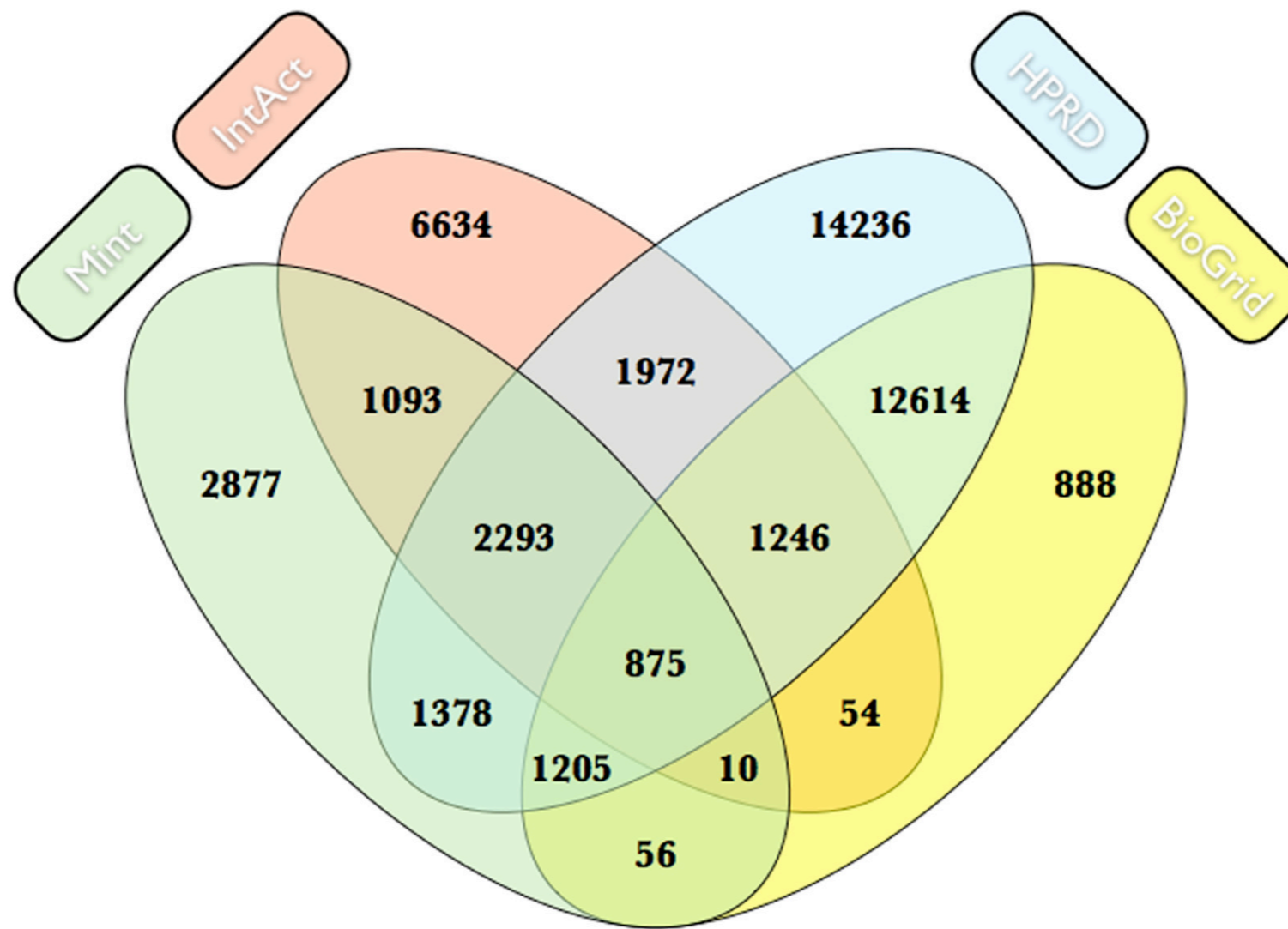
DATA FUSION



DATA FUSION



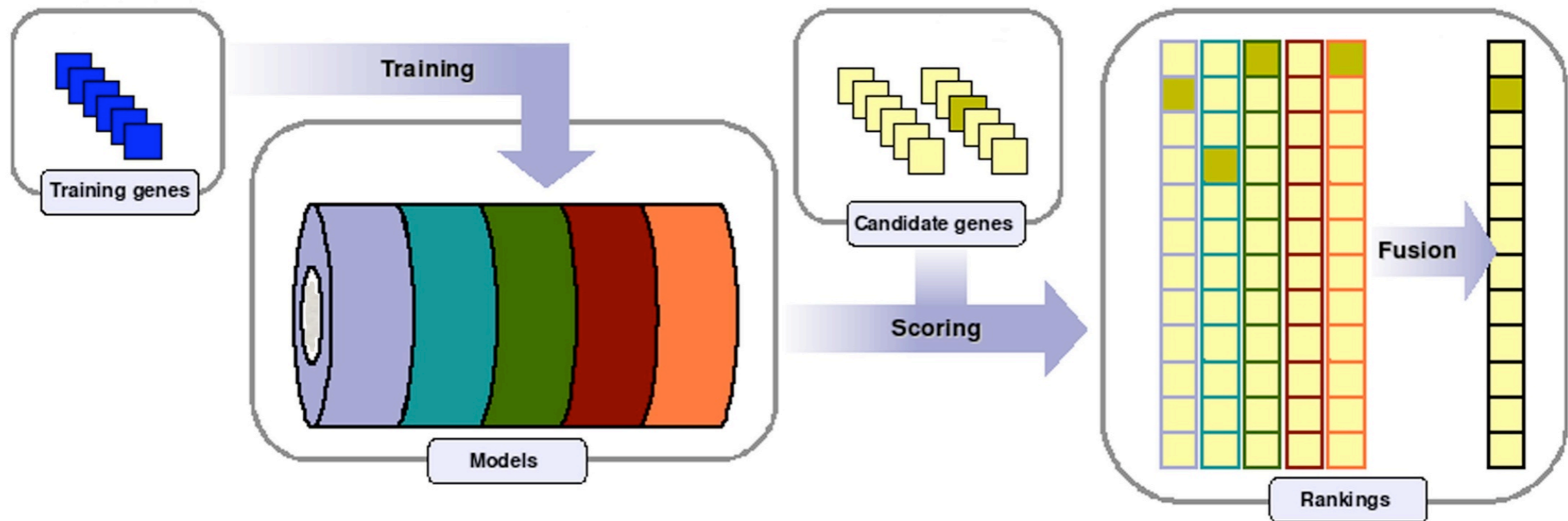
DATA FUSION



**ORDER STATISTICS BASED
DATA FUSION**

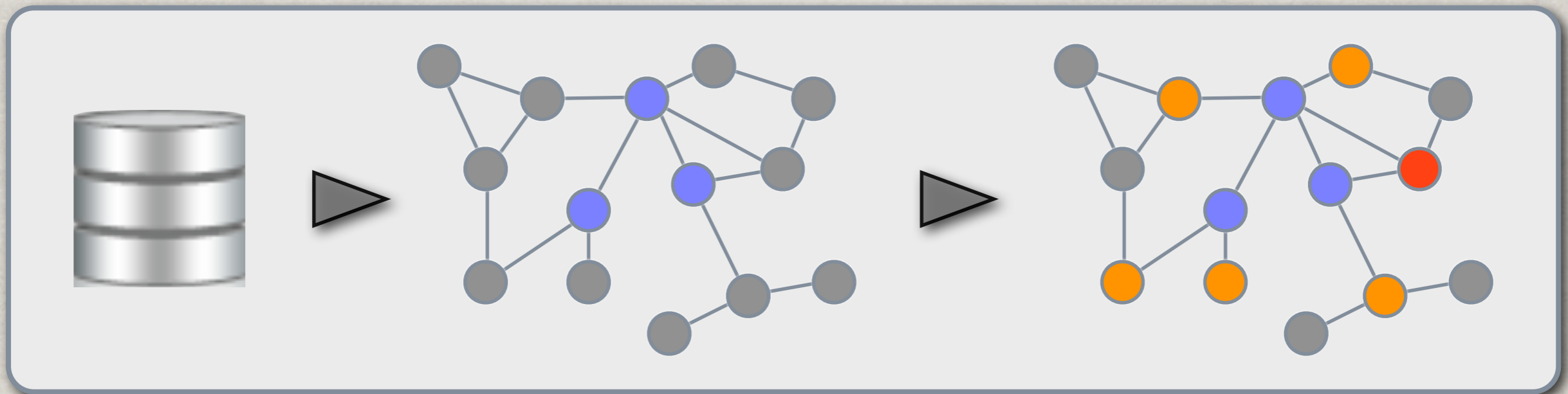


ALGORITHM



Tranchevent *et al.*, Nucleic Acids Research (2008)

SIMILARITY

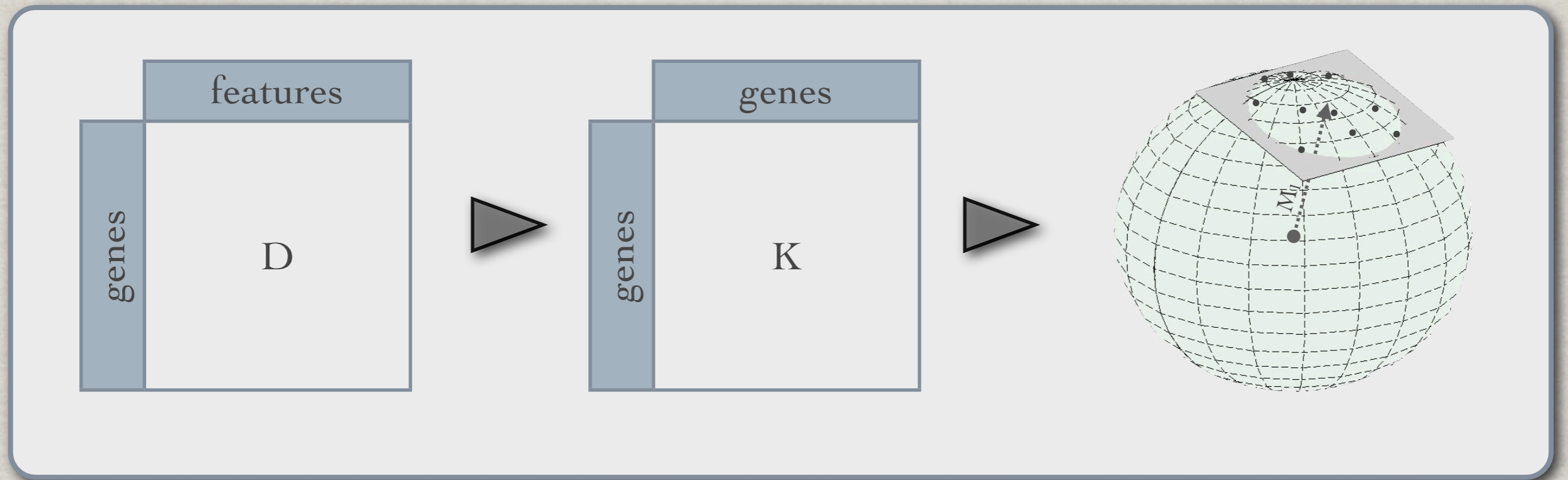


Neighbors of known disease genes are promising candidates

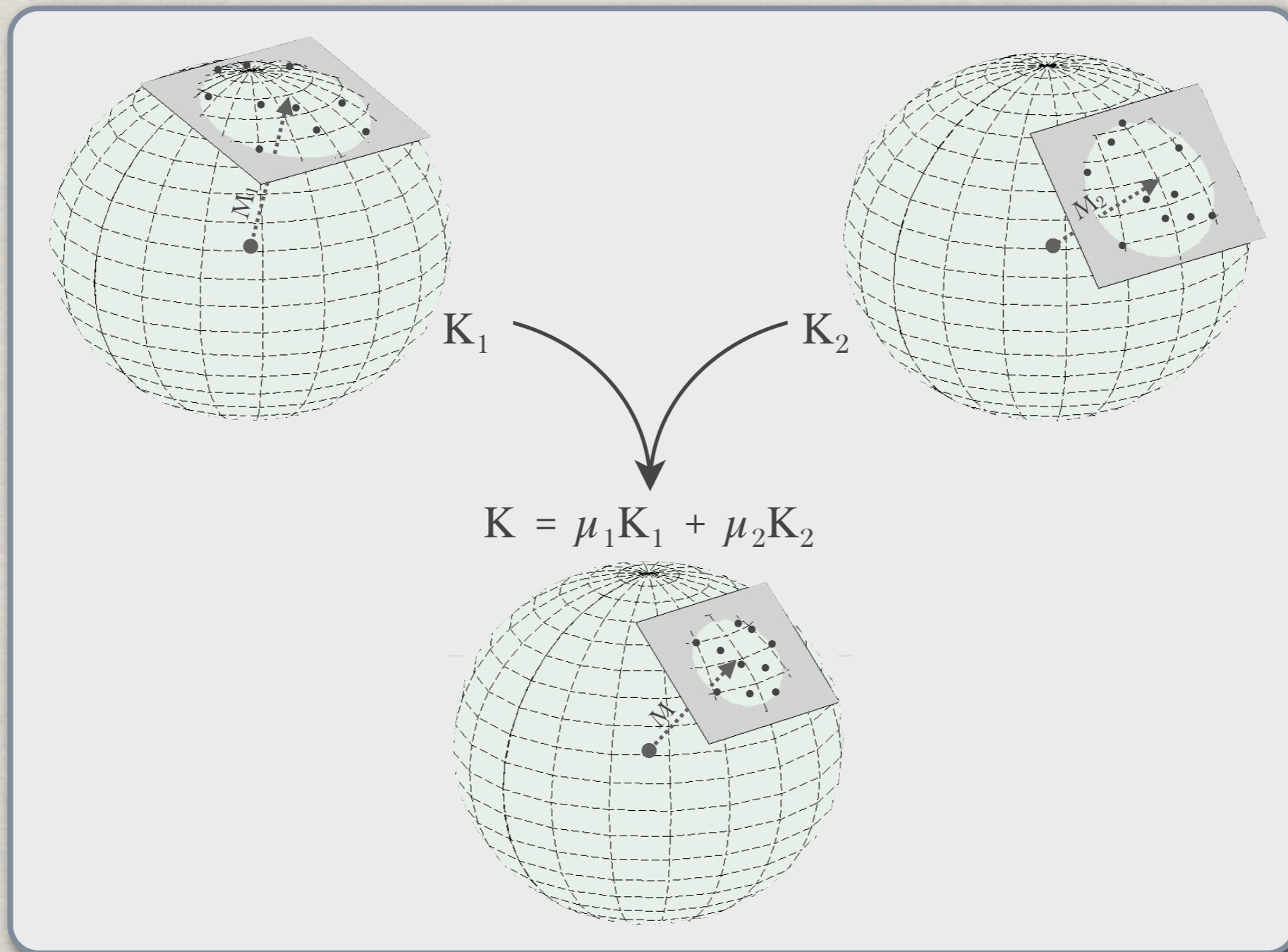
KERNEL BASED DATA FUSION



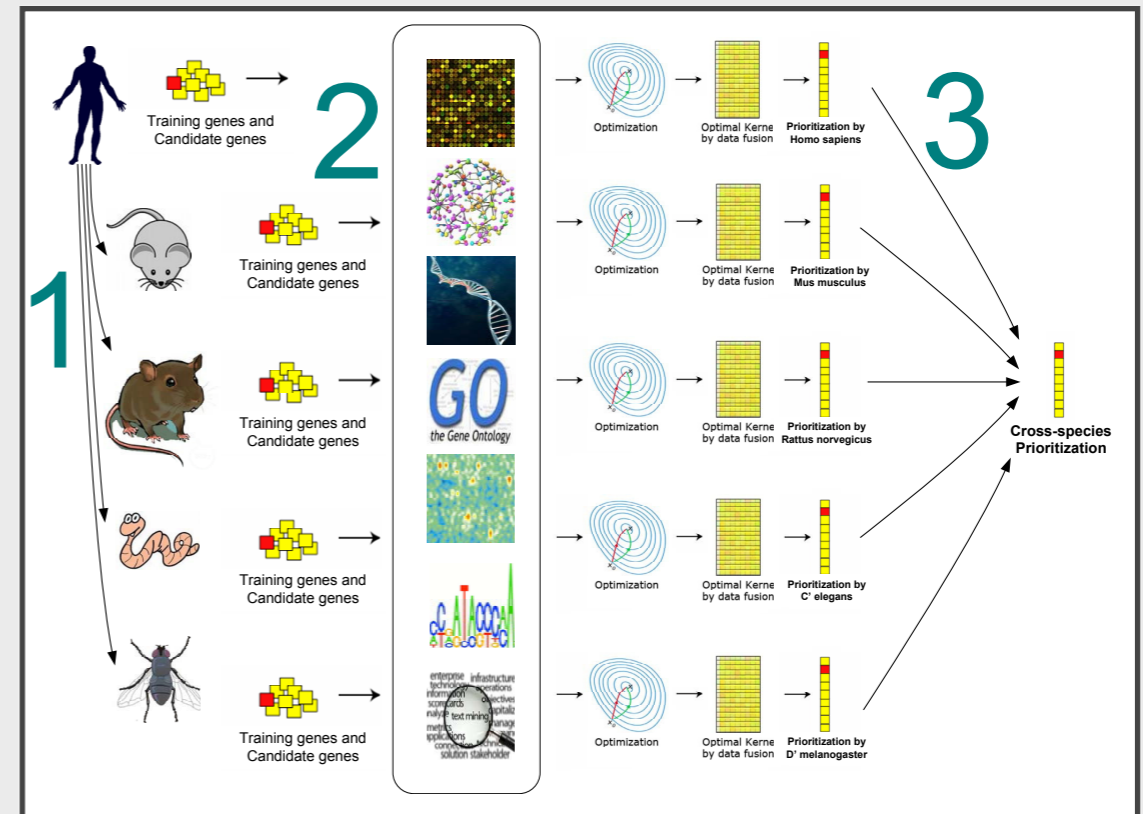
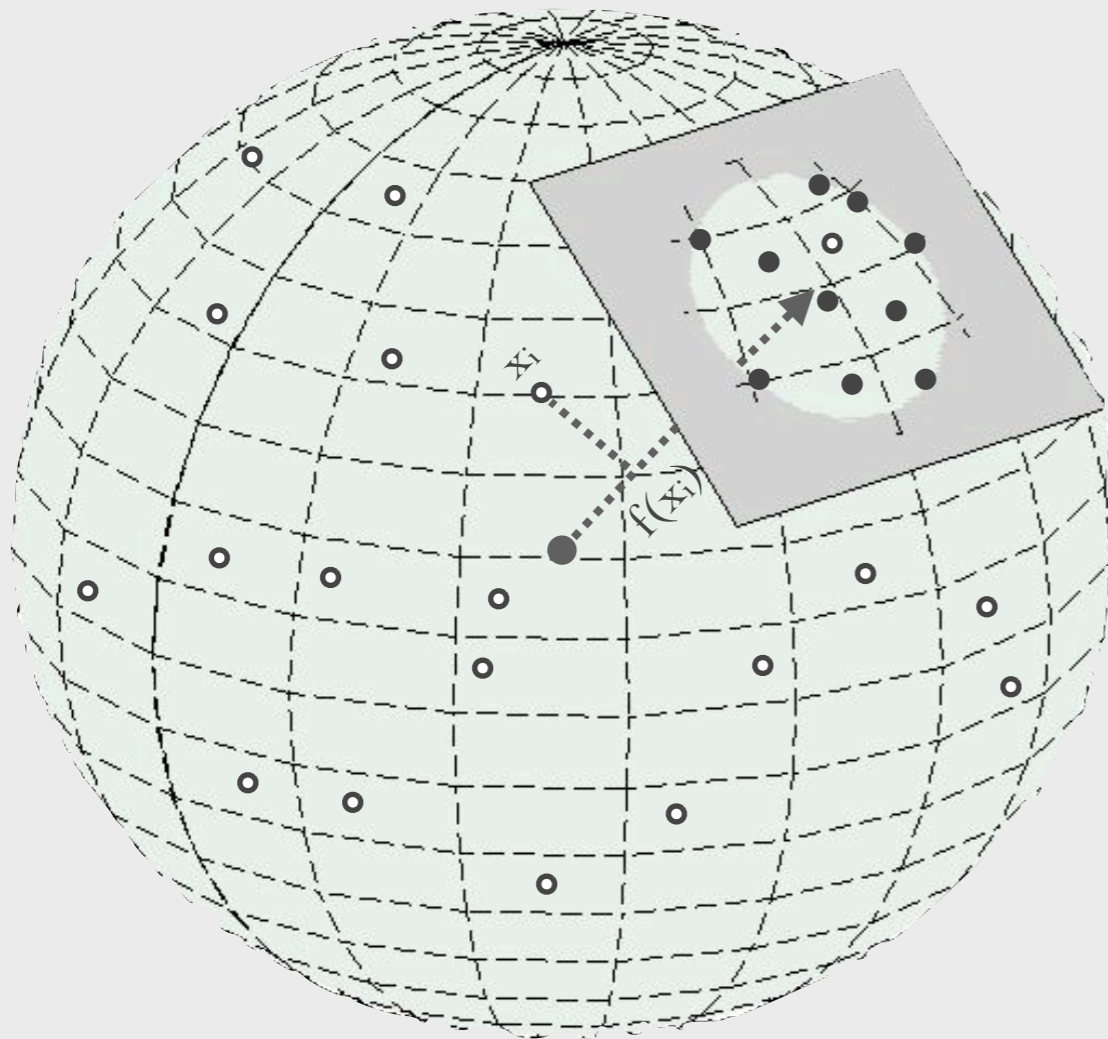
ALGORITHM



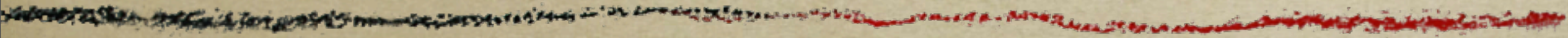
ALGORITHM



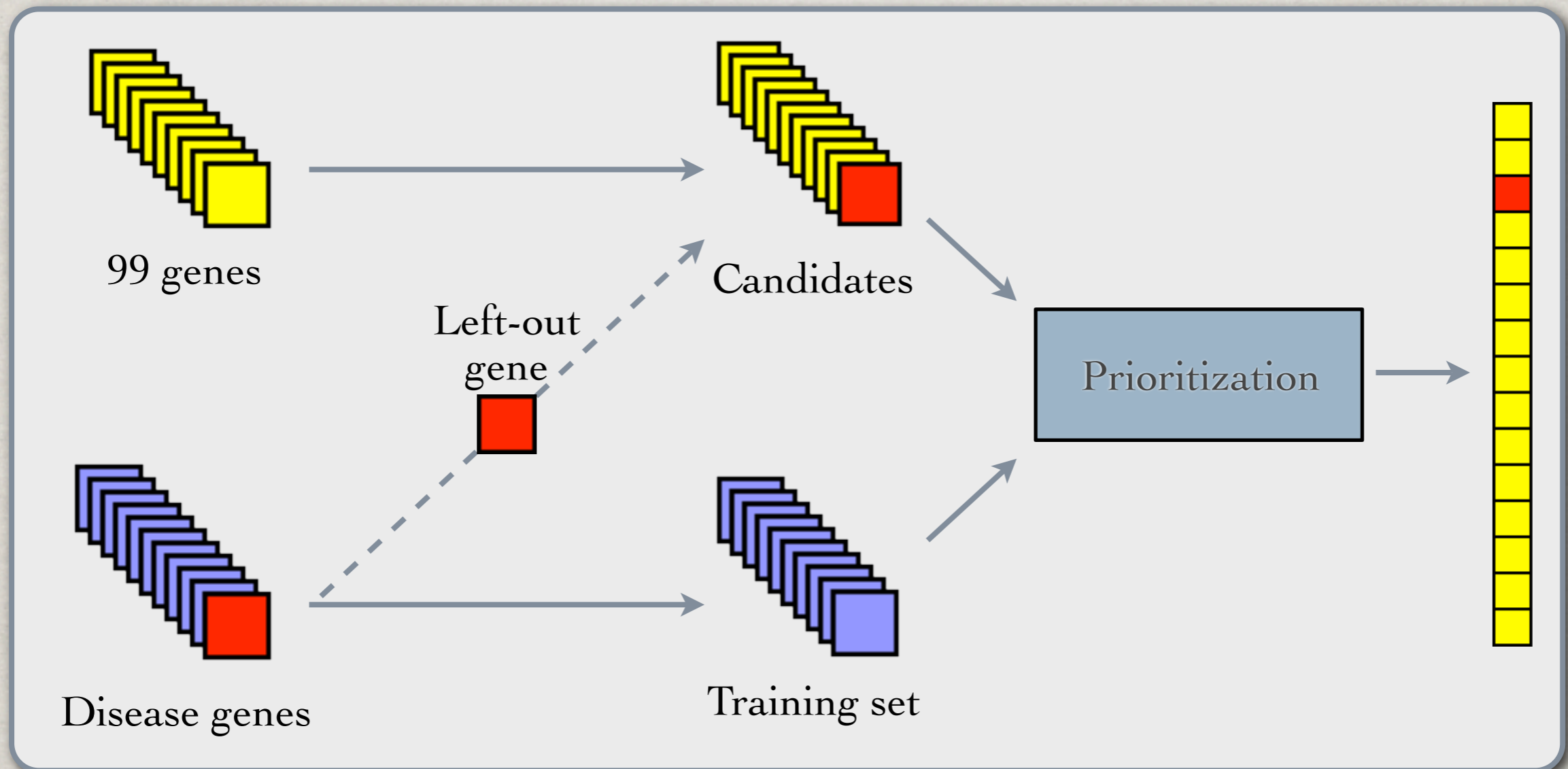
ALGORITHM



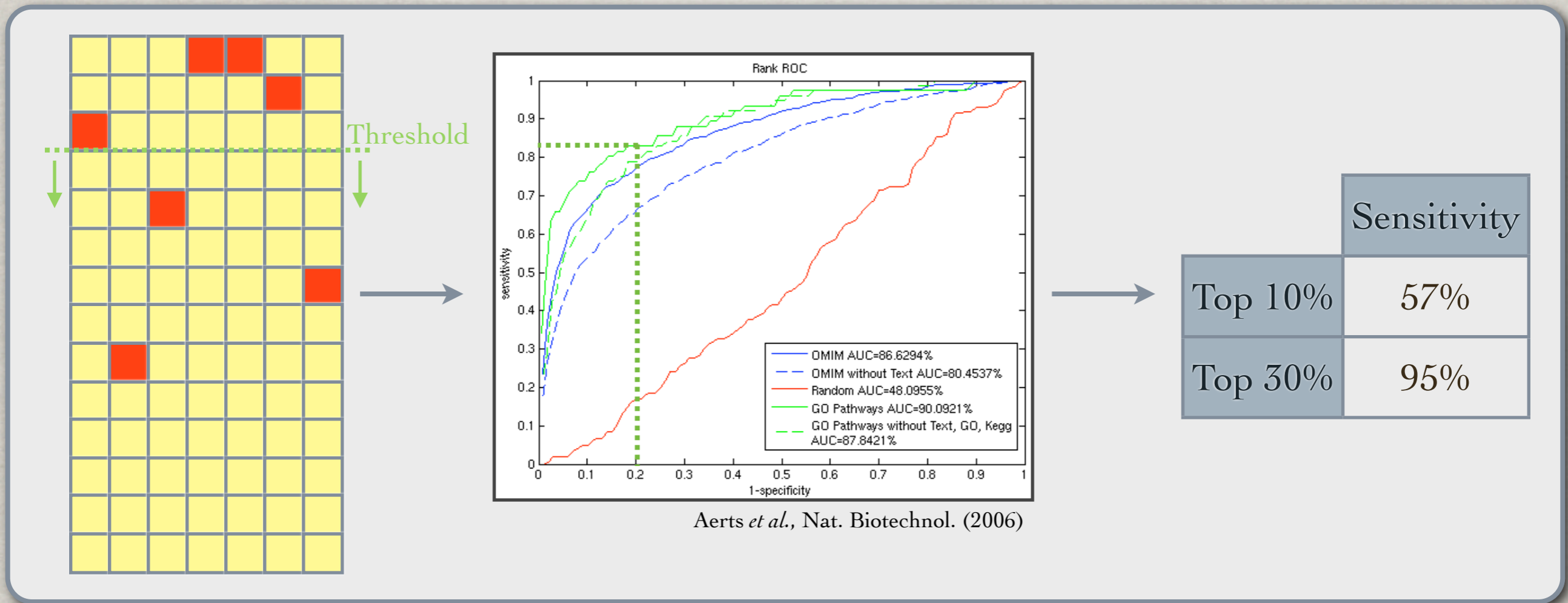
BENCHMARKING



PROCEDURE



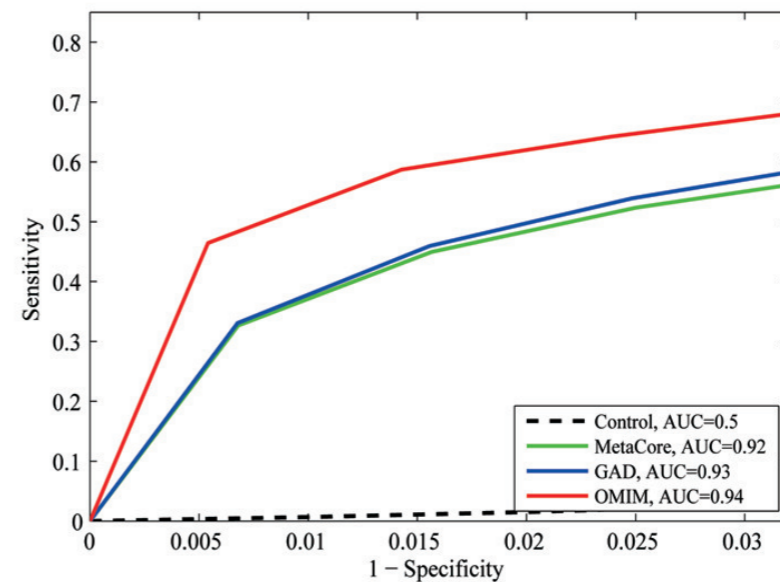
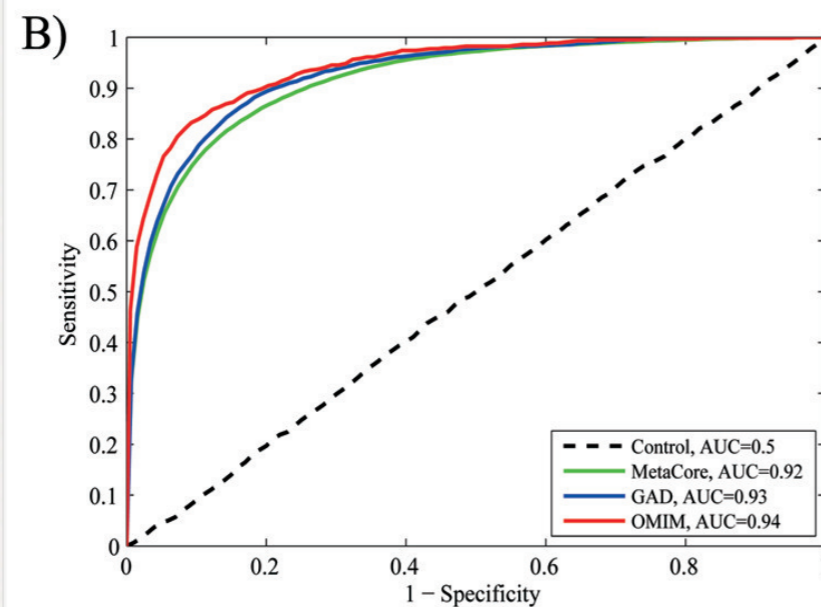
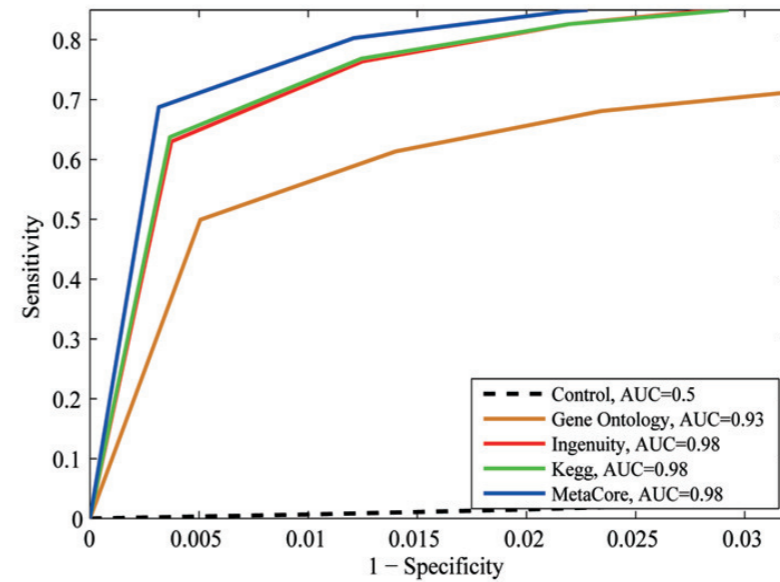
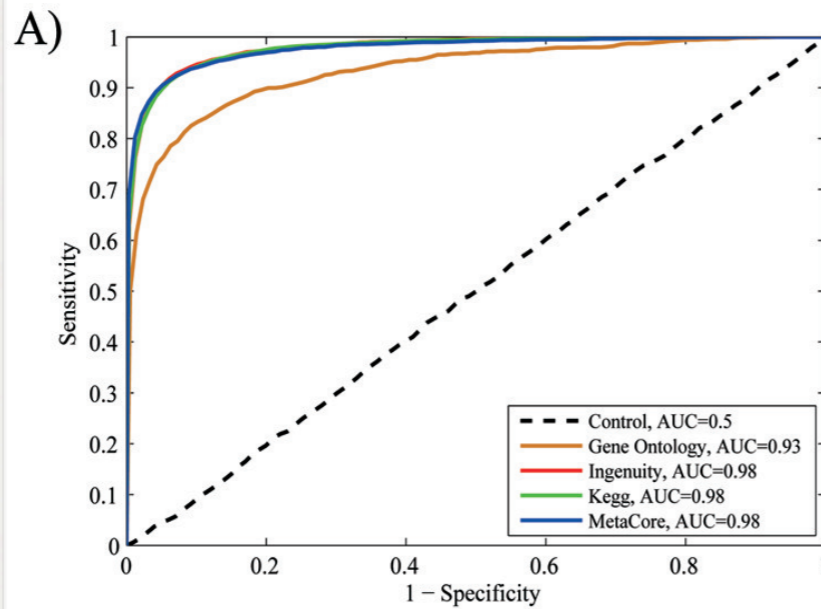
PROCEDURE



DATASETS

	OMIM	GO I	MetaCore diseases	GAD	MetaCore pathways	Ingenuity pathways	GO II
Nb sets	29	3	833	142	454	94	37
Nb genes	621	75	12699	3310	10053	2947	1216

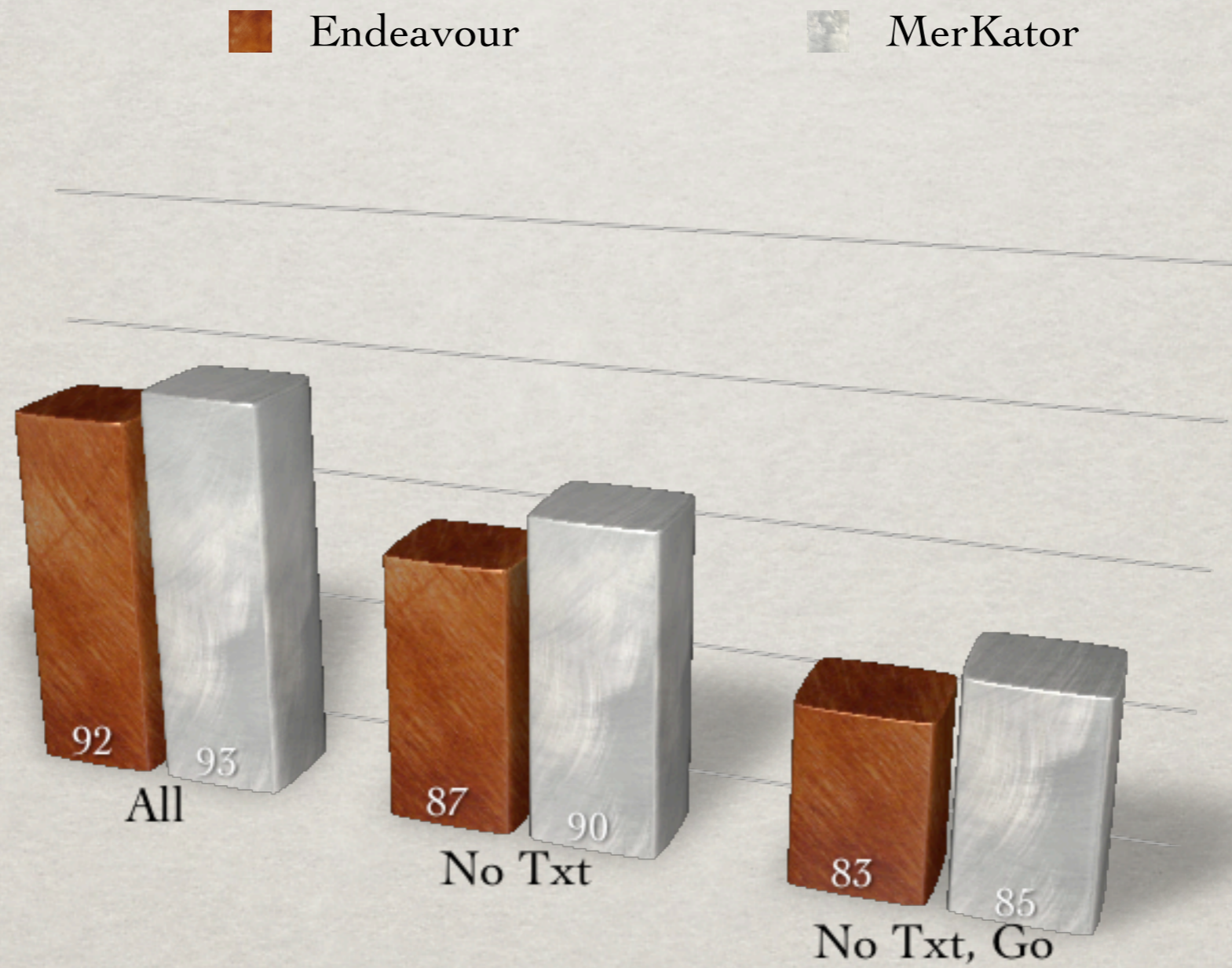
ENDEAVOUR: RESULTS



ENDEAVOUR: RESULTS

	Random	OMIM	MetaCore diseases	GAD	MetaCore pathways	Ingenuity	GO II
Sensitivity at 1%	~1%	46%	33%	33%	64%	69%	56%
Sensitivity at 10%	10%	83%	75%	77%	94%	94%	89%
Sensitivity at 30%	30%	95%	92%	93%	98%	98%	97%

MERKATOR: RESULTS



ENDEAVOUR: LITERATURE

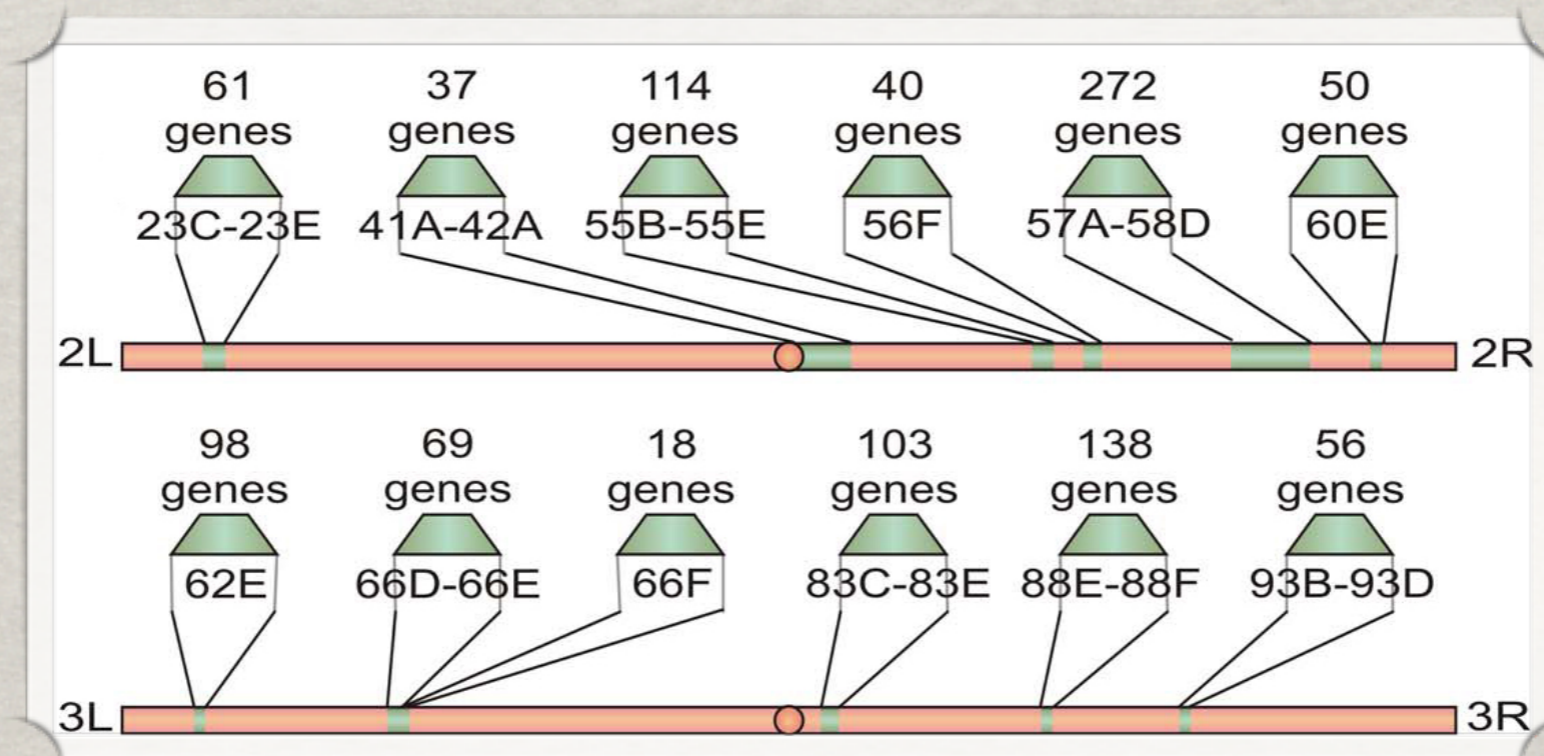
	Literature (2008)	Literature (2010)
Number of genes	32	43
Median	9,50%	11,21%
Sensitivity at 30%	90,63%	90,70%

VALIDATION



GENETIC SCREEN

- ✿ 180 deficiency lines on chromosomes 2 and 3
- ✿ 12 *ato* specific positive loci (1056 genes)

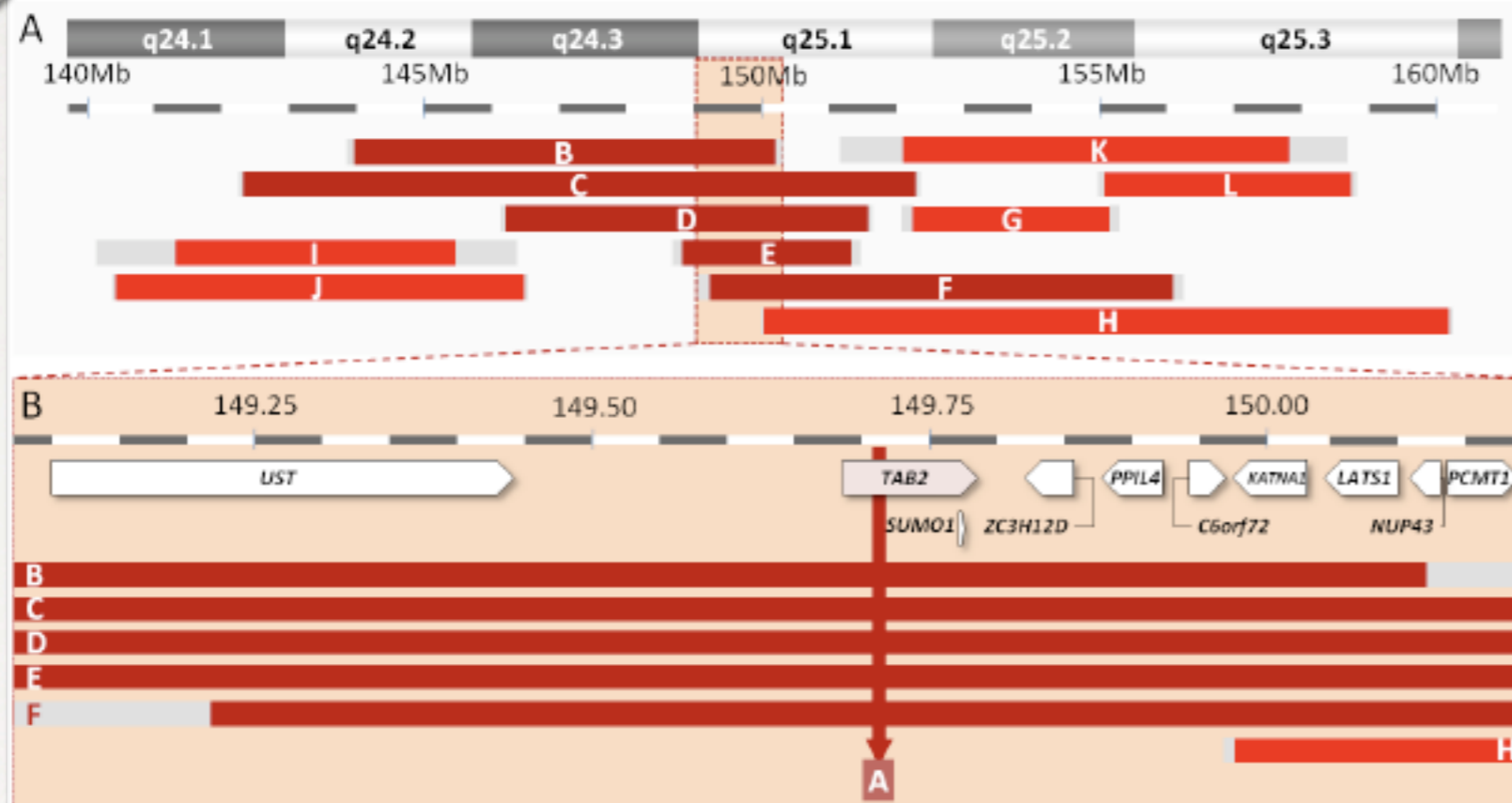


Aerts *et al.*, Plos Genetics (2009)

GENETIC SCREEN

Name	Rank ratio on locus	Rank ratio on chromosome	Phenotype
cas	1,94%	2,00%	-33,60%
dom	5,15%	7,90%	-31,40%
Egfr	0,37%	0,03%	-51,30%
fj	2,63%	1,20%	-50,00%
lilli	8,20%	4,70%	-35,40%
mus209	2,50%	9,20%	-36,70%
ppan	14,29%	10,70%	-33,60%
sbb	7,02%	3,90%	-33,60%
shg	0,74%	0,60%	-66,40%
smg	5,56%	4,60%	-100%
toc	14,74%	15,90%	-33,60%
zip	2,00%	2,30%	-38,10%

CHD



Thienpont *et al.*, Am. J. Hum. Genet. (2010)

B

rank	Neural Crest	1 st Heartfield	2 nd Heartfield	Valve formation	Vascularisation	Combination
1	TAB2	OPRM1	SOD2	IGF2R	TAB2	TAB2
2	ESR1	ZBTB2	TAB2	TAB2	ESR1	OPRM1
3	WTAP	TAB2	ARID1B	MTRF1L	OPRM1	ESR1
4	SNX9	SOD2	PPP1 R14	VIL2	IGF2R	SOD2
5	IGF2R	PPP1 R14	OPRM1	SYNE1	PPP1 R14	IGF2R
6	ZBTB2	ESR1	ZBTB2	OPRM1	LATS1	ZBTB2

Thienpont *et al.*, Am. J. Hum. Genet. (2010)

**CONCLUSIONS
AND
PERSPECTIVES**



CONCLUSIONS

- ✻ Two prioritization methods
- ✻ Benchmarked
- ✻ Experimentally validated

PERSPECTIVES

- ✻ Improved data warehouse
- ✻ Automatic training set construction
- ✻ Bio-entities prioritization
- ✻ Improved results visualization

PERSPECTIVES

- ✱ Chemical/pharmaceutical focus
- ✱ Clustering/classification/prioritization platform
- ✱ Feature selection before kernel computation
- ✱ More experimental validations