# Using public RNA-sequence data to predict rare disease genes







n F



Patrick Deelen

24-02-2021

111

# Our current knowledge of genes is biased

# THE TOP 10

The ten most studied genes of all time are described in more than 40,000 papers.



"Out of the 20,000 or so protein-coding genes in the human genome, just 100 account for more than one-quarter of the papers"

Elie Dolgin, Nature 2015

# Using 30,000 RNA-seq samples to predict gene function



# Prediction of gene functions



# Human Phenotype Ontology (HPO)



Test performance of predictions

## 8,657 Pathways / HPO Terms



Predicted pathways and disease symptoms for all genes

-5

Gene prioritization Z-score

15

56,435 genes

# Using GADO in a diagnostic setting

### Patient specific input:

- Symptoms (HPO terms)
- Set of ± 250 genes that have potential disease causing variant

### GADO:

→ Prioritization of genes predicted to cause the symptoms of a patient

### Patient specific output:

Input set ranked based on the genes that are most likely to cause symptoms





### Predict gene functions. Discover potential disease genes. Search here or paste a list of multiple genes (Ensembl IDs or HGNC symbols) -For example: SMIM1, Interferon signaling, Migraine, Autism TOOLS HPO gene prioritisation Function enrichment Prioritize genes based on one or Predict which pathways are enriched multiple HPO phenotypes. for a set of genes. ightarrow continue $\rightarrow$ continue N. 候 umcg / rijksuniversiteit groningen **CLEVER°FRANKE** © 2017 Department of Genetics, University Medical Center Groningen www.genenetwork.nl

### GENE ♦ NETWORK

### HPO GENE PRIORITIZATION

Select			-	
TERM	ID	REM	DVE	
Abnormality of the face	HP:0000271	Х	^	
Neoplasm of the endocrine system	HP:0100568	Х		
Abnormality of the anterior pituitary	HP:0011747	Х		
Intellectual disability	HP:0001249	Х		
• OPTIONAL: filter output on candida	ate genes			
OPTIONAL: filter output on candida  MI-CYB  CHOOSE A FILE	ate genes			

 $\leftarrow$  GO BACK

候 umcg



 $\mathbf{w}$ 

PHENOT	YPE		ANNOTATED GENES	HPO-TERM				
Abnorm	ality of the fa	ice	1948	HP:0000271				
Neoplas	sm of the end	locrine system	88	HP:0100568				
Abnorm	ality of the a	nterior pituitary	186	HP:0011747				
Intellec	tual disability		1289	HP:0001249				
<b>enes r</b> 9orf47,	not found LOC1014482	02, C9orf172, C15orf	f38-AP3S2, C22orf46	Kleefstra S	yndrome			
enes f	ound							
he 100	prioritized sel	ected genes for the o	combination of these 4 phenotypes:				0	•
					-	32 <sup>17</sup>	Se N	141 -0124
	RANK	GENE	Z-SCORE	NETW	DRK 48 <sup>900</sup>	HP:010	HEIDDY	HPIDDL
			0.001		UIL	1.0	5.5	Į.
	2	EHMT1	2.961		2.6	0.6	1.3	1.5
-	0		2.070		1.0	1 1	1.1	1
÷.,	4		2.457	*** •**	1.7	1.1	1.1	1
2.1	5	SLU2DAD	2.345		0.8	0.7	-0.8	4.1
÷	0	11110	2.20		1.3	0.7	-U.I 1 0	2.0
	, D	NIID133	2.220		-0.7	1.0	1.9	21
÷.	q	FIE/JENIE1	2.131		11	0.7	0.0	3
1.1	10	DNA2	1 95	•	0.1	15	0.5	17
÷	11	ATP50	1.55	•.	0.5	0.7	17	0.9
	12	MAML 2	1.865	•*	1	0.9	0.8	1.1
	13	DMXL2	1.847	-*	0.9	1.1	0	1.7
	14	GRK5	1.823	-	0.4	1.4	1.5	0.3
	15	TP73	1.805	*	1.4	1.5	0.5	0.2
	16	PALM	1.782	-*	1.3	1.5	1	-0.2
	17	ו סראס	1 755	**	0.8	-0.1	11	17

# Unsolved case: OBSCN

GENE NETV	: VORK v2.0	Search here or paste a list of multiple gen	nes (En 🔻 HOM	ie faq api
PHENOTY	PE		ANNOTATED GENES	HPO-TERM
Dilated car	rdiomyopathy		99	HP:0001644
<b>Genes not</b> MFSD14B, C	t <b>found</b> DISP3			c HU
Gene prio RANK	GENE	Z-SCORE	NETWORK	HP:00016
Gene prio RANK	GENE CACNA1S	Z-SCORE 7.8	NETWORK	499.00016
Gene prior RANK	GENE CACNA1S RPL3L	Z-SCORE 7.8 7.1	NETWORK	499,00016 7.8 7.1
Gene prior RANK	GENE CACNA1S RPL3L OBSCN	Z-SCORE 7.8 7.1 6	NETWORK	499,00016 7.8 7.1 6
Gene prior RANK 1 2 3 4	ritization GENE CACNA1S RPL3L OBSCN CALU	Z-SCORE 7.8 7.1 6 4.3	NETWORK	4,3
Gene prior RANK 1 2 3 4 5	ritization GENE CACNA1S RPL3L OBSCN CALU PLXNA1	Z-SCORE 7.8 7.1 6 4.3 3.6	NETWORK	4.3 3.6
Gene prior RANK	ritization GENE CACNA1S RPL3L OBSCN CALU PLXNA1 OPN4	Z-SCORE 7.8 7.1 6 4.3 4.3 3.6 3.5	NETWORK	4.3 3.6 3.5
Gene prio RANK 1 2 3 4 5 6 7	ritization GENE CACNA1S RPL3L OBSCN CALU CALU PLXNA1 OPN4 COL18A1	Z-SCORE 7.8 7.1 6 4.3 4.3 3.6 3.5 3.2	NETWORK	HP:001b 7.8 7.1 6 4.3 3.6 3.5 3.2
Gene prior RANK 1 2 3 4 5 6 7 8	ritization GENE CACNA1S RPL3L OBSCN CALU CALU PLXNA1 OPN4 COL18A1 MUL1	Z-SCORE 7.8 7.1 6 4.3 4.3 3.6 3.5 3.5 3.2 2.9	NETWORK ペ ペ ペ ペ ペ ペ ペ ペ ペ ペ ペ ペ ペ	4.3 3.6 3.2 2.9
Gene prior RANK 1 2 3 4 5 6 7 8 9	ritization GENE CACNA1S RPL3L OBSCN CALU CALU PLXNA1 OPN4 COL18A1 MUL1 ALPK2	Z-SCORE 7.8 7.1 6 4.3 4.3 3.6 3.5 3.5 3.2 2.9 2.7	NETWORK	4.3 3.6 3.2 2.9 2.7

# Unsolved case: OBSCN



# OBSCN predicted functions

GENE NETWORK v2.0 Search here or paste a list of multiple genes (Ensemble	IDs or HGNC symbols)	- Home	FAQ API
OBSCN obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF Gene predictability score: 0.66 (1)		ct P	nromosome 1 rotein coding
SHOW       PATHWAYS & PHENOTYPES       CO-REGULATED GENES       TISSUE         SELECT DATABASE       REACTOME       HUMAN PHENOTYPE ONTOLOGY       GO biological press         GO CELLULAR COMPONENT       KEG PATHWAY       TERM	ES PROCESS GO MOLECULAR P-VALLIE DIR	RECTION ANNOTATE	) NETWORK
cardiac muscle contraction	8.7 × 10 <sup>-9</sup>	<b>_ _</b>	
akalatal musala contraction	-		
Skeletal muscle contraction	1.5 × 10 <sup>-6</sup>	· · · · ·	*
muscle filament sliding	1.5 × 10 <sup>-6</sup> 2.2 × 10 <sup>-6</sup>	▲ <del>–</del>	-х -х
muscle filament sliding sarcomere organization	1.5 × 10 <sup>-6</sup> 2.2 × 10 <sup>-6</sup> 7.7 × 10 <sup>-6</sup>		ス イ イ
muscle filament sliding sarcomere organization striated muscle contraction	1.5 × 10 <sup>-6</sup> 2.2 × 10 <sup>-6</sup> 7.7 × 10 <sup>-6</sup> 4.7 × 10 <sup>-5</sup>		ス イ イ イ
skeletal muscle contraction         muscle filament sliding         sarcomere organization         striated muscle contraction         striated muscle myosin thick filament assembly	$1.5 \times 10^{-6}$ $2.2 \times 10^{-6}$ $7.7 \times 10^{-6}$ $4.7 \times 10^{-5}$ $4.4 \times 10^{-4}$		メ メ メ メ 、

# Dilated cardiomyopathy genes + OBSCN



# Pathway enrichment analysis



# Using GADO in practice

### Cohort of 83 solved cases

- Each case has on average 56 variants in known disease-causing genes
- In 42% of cases the causal variant is in the top 3
- Cohort of 61 cases not solved by routine diagnostics
  - We found a strong candidate gene for 10 cases
  - Due to a stringent cut-off on average 2.9 genes per cases were looked at in detail















GWAS summary statistics give each gene a p-value

Co-regulation captures relatiosnhips between genes in a tissue

Integration prioritises core genes strongly co-regulated with genes inside GWAS loci



# Predicted key-genes often loss of function intolerance

Prioritizing core genes for 25 traits and yielded 158 significant key genes



# IBD key gene prioritization



### NFKB1 is co-regulated with *cis* and *trans* genes



# IBD key gene prioritization



Rare disease gene prediction using IBD key gene scores

Rare disease symptom (HPO)	AUC
Abnormal delayed hypersensitivity skin test	0.93
Recurrent candida infections	0.93
Recurrent staphylococcal infections	0.91
Lymphocytosis	0.90
Chronic mucocutaneous candidiasis	0.89
Recurrent cutaneous fungal infections	0.89
Autoimmune hemolytic anemia	0.88
Antiphospholipid antibody positivity	0.87
B lymphocytopenia	0.86
Eczematoid dermatitis	0.85

Strongly depleted for loss of function alleles

Enriched for known rare disease genes

Can be used to predict new rare disease genes



Large effect

# KidneyNetwork



# KidneyNetwork

A) Prediction accuracy of 99 kidney related HPO-terms









# KidneyNetwork



#### KidneyNetwork incorporated in the GADO method on

# Future improvements – directed networks



# Growth of public RNA-seq

Number of publicly available human RNA-seq samples in the European Nucleotide Archive



# Acknowledgements

Annique Claringbould Olivier Bakker Floranne Boulogne Laura R. Claus Floor Schukking Sophie Mulcahy Symmons Sipko van Dam Johanna C. Herkert Juha M. Karjalainen Harm Brugge Kristin M.Abbott Cleo C. van Diemen Paul A. van der Zwaag Erica H. Gerkes **Evelien Zonneveld-Huijssoon** Jelkje J. Boer-Bergsma Pytrik Folkertsma

Tessa Gillett K. Joeri van der Velde Roan Kanninga Peter C. van den Akker Sabrina Z. Jan Edgar T. Hoorntje P te Rijdt Wouter Yvonne |. Vos Jan D.H. Jongbloed Conny M.A. van Ravenswaaij-Arts Richard Sinke Birgit Sikkema-Raddatz Wilhelmina S. Kerstjens-Frederikse Morris A. Swertz Nine V.A.M. Knoers Albertien M. van Eerde Lude Franke







