Disease variant prioritisation and model discovery through cross-species phenotype analysis

Damian Smedley, Sanger Institute

monarch INITIATIVE

IMPC MPI2

Standard exome analysis

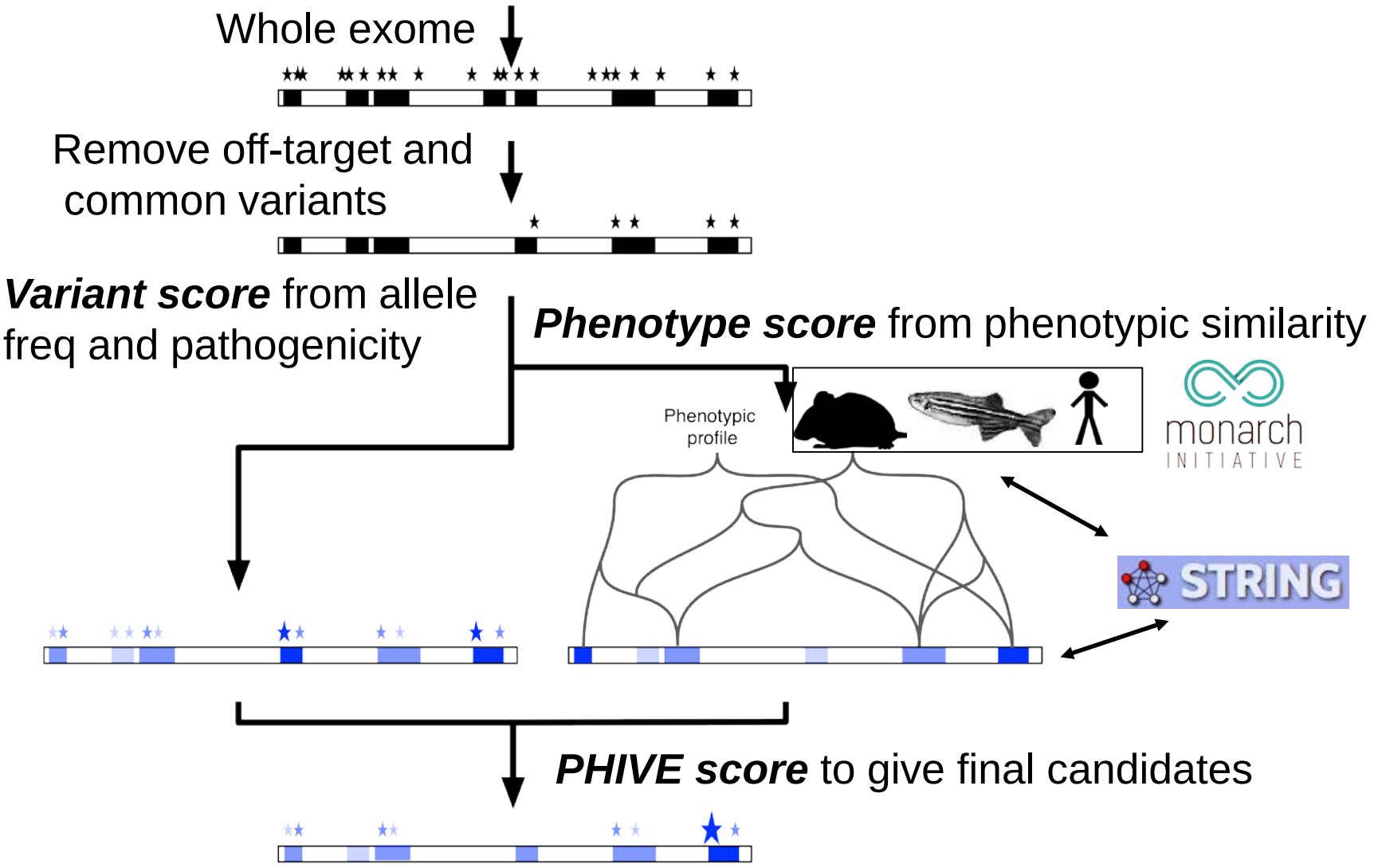
Whole exome

Remove off-target and common variants

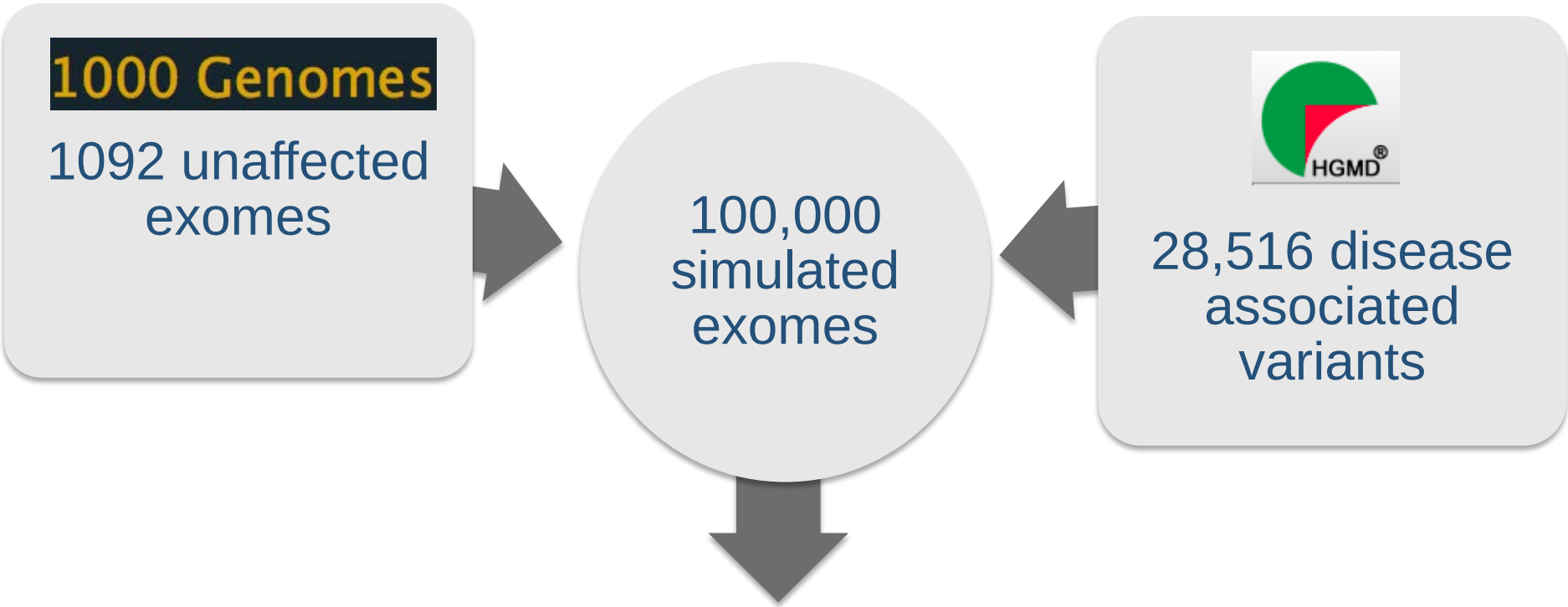
Prioritize based on allele frequency and pathogenicity

Linkage data
Multiple affected individuals
De novos from trios
Prior knowledge of affected pathways

Exomiser

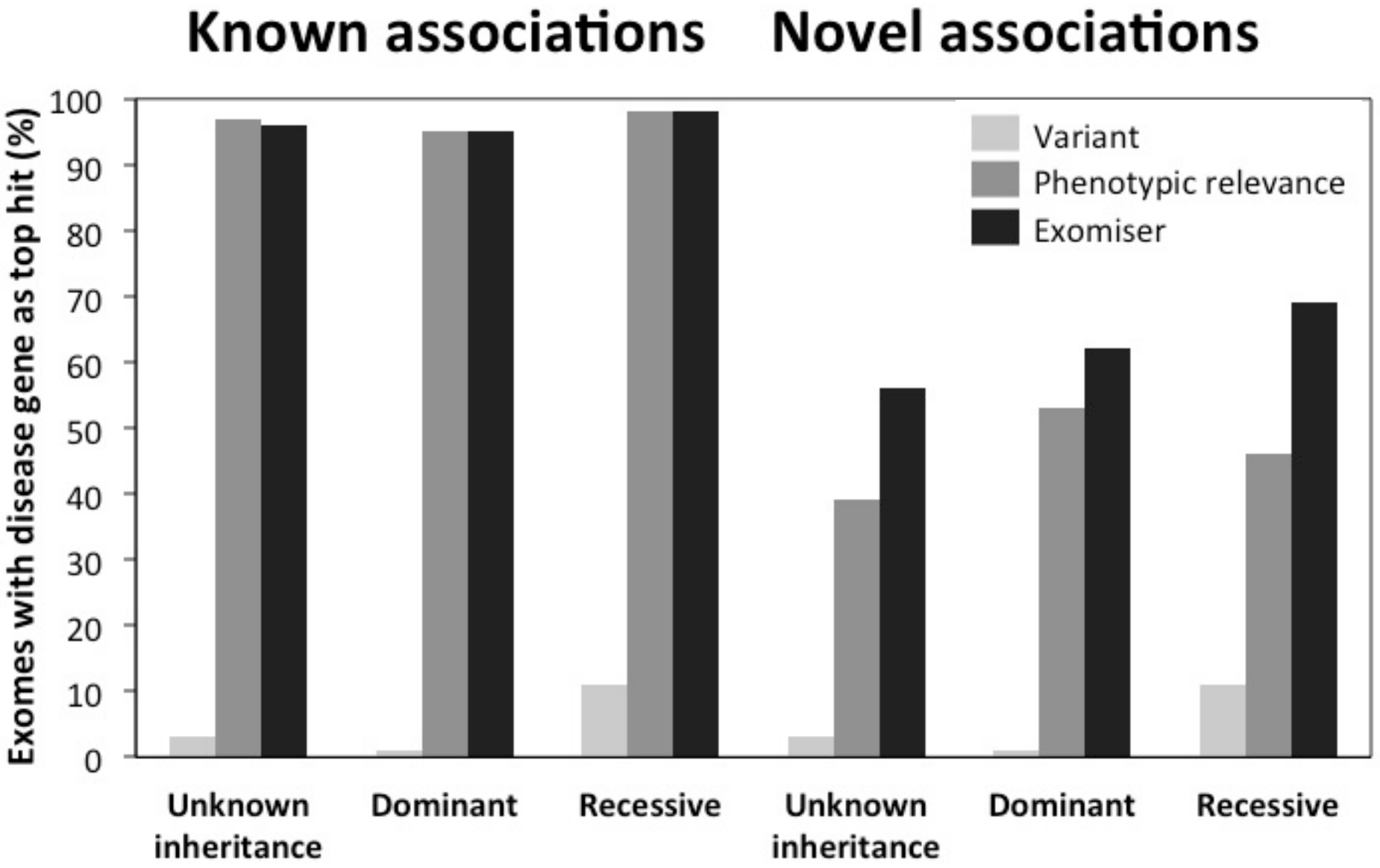

Whole exome
Remove off-target and common variants
Variant score from allele freq and pathogenicity
Phenotype score from phenotypic similarity
Phenotypic profile
monarch INITIATIVE
STRING
PHIVE score to give final candidates

Benchmarking

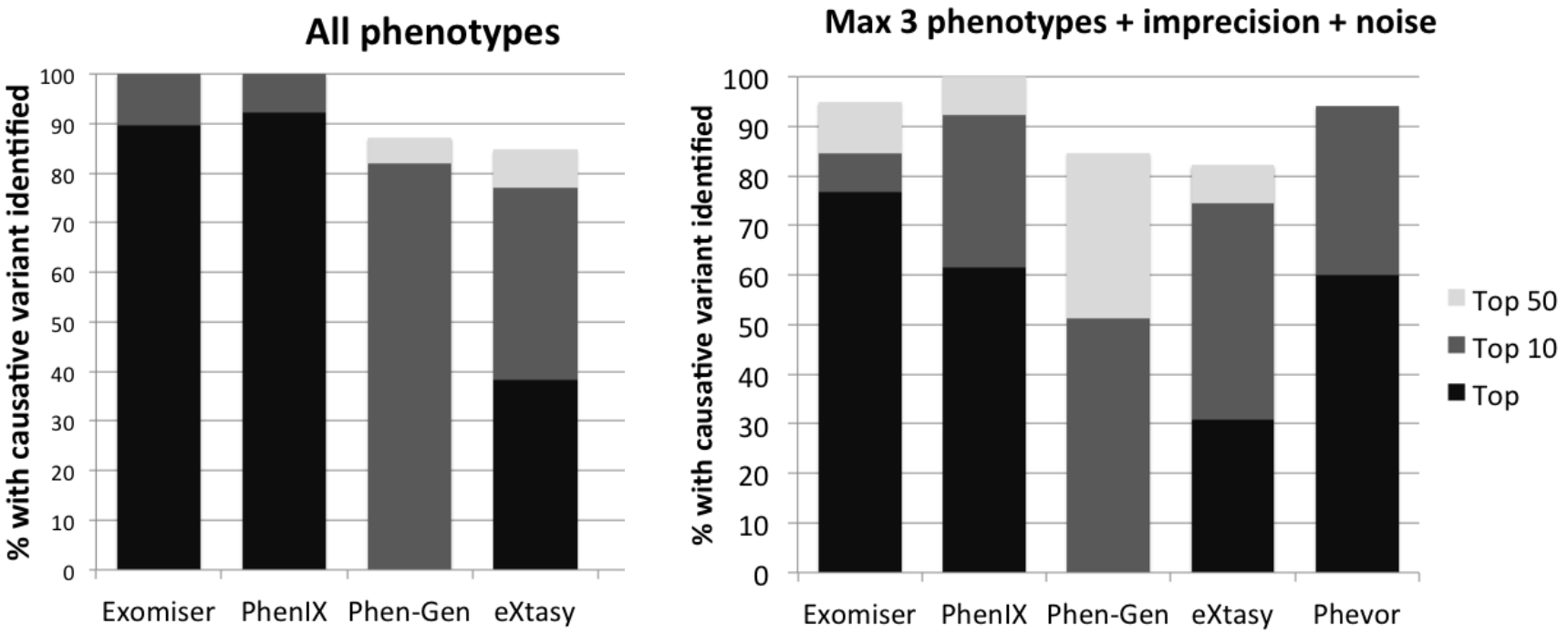


- Annotate variants
- Remove off-target, syn and common(>1%) variants (plus optional inheritance model filtering)
- Prioritize based on combined score

Variant and phenotype data synergistically identify causative variant

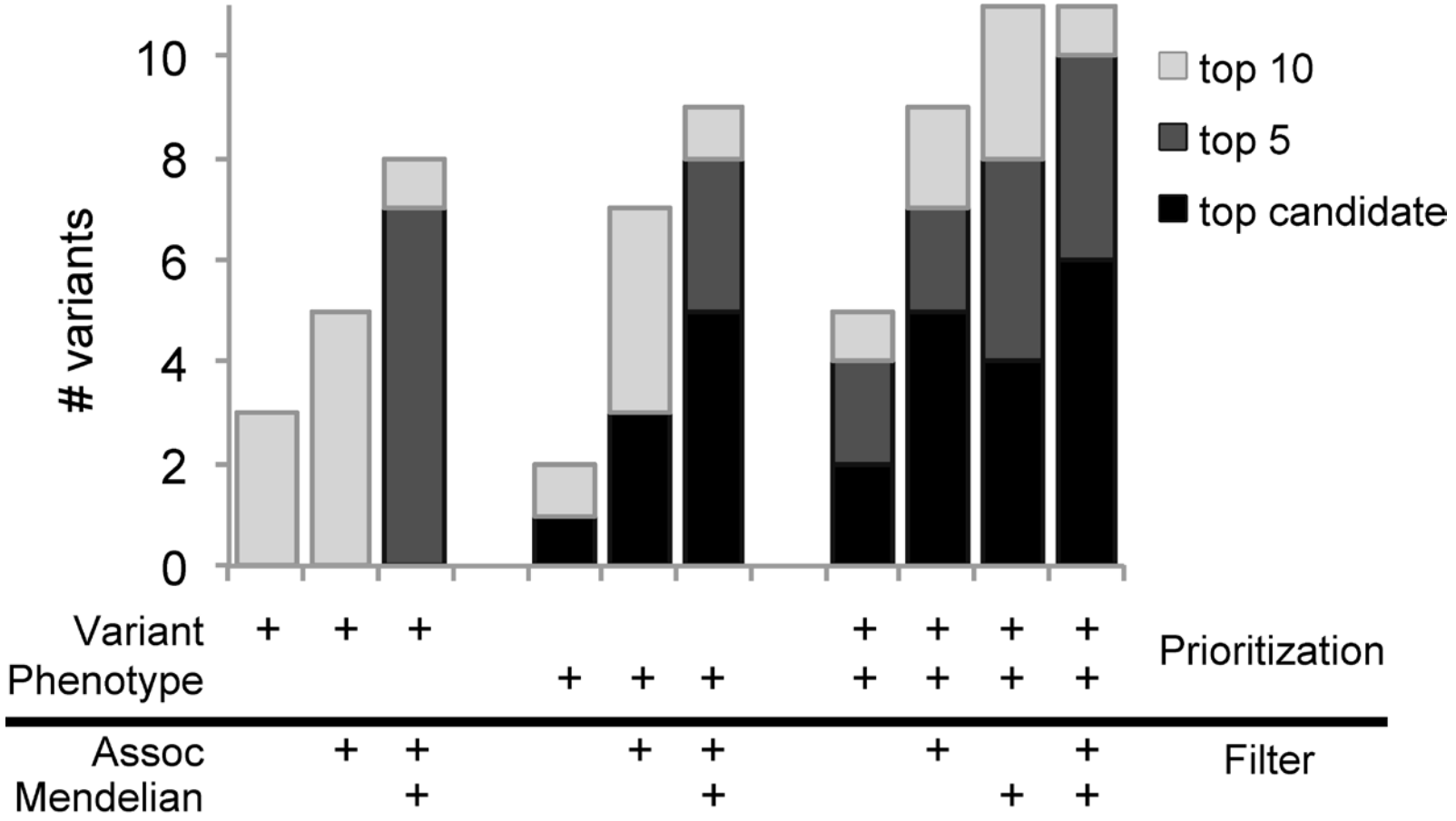


Comparison to other phenotype-based variant analysis software



Smedley D & Robinson PN. Genome Medicine 2015, 7:81

NIH Undiagnosed Disease Program



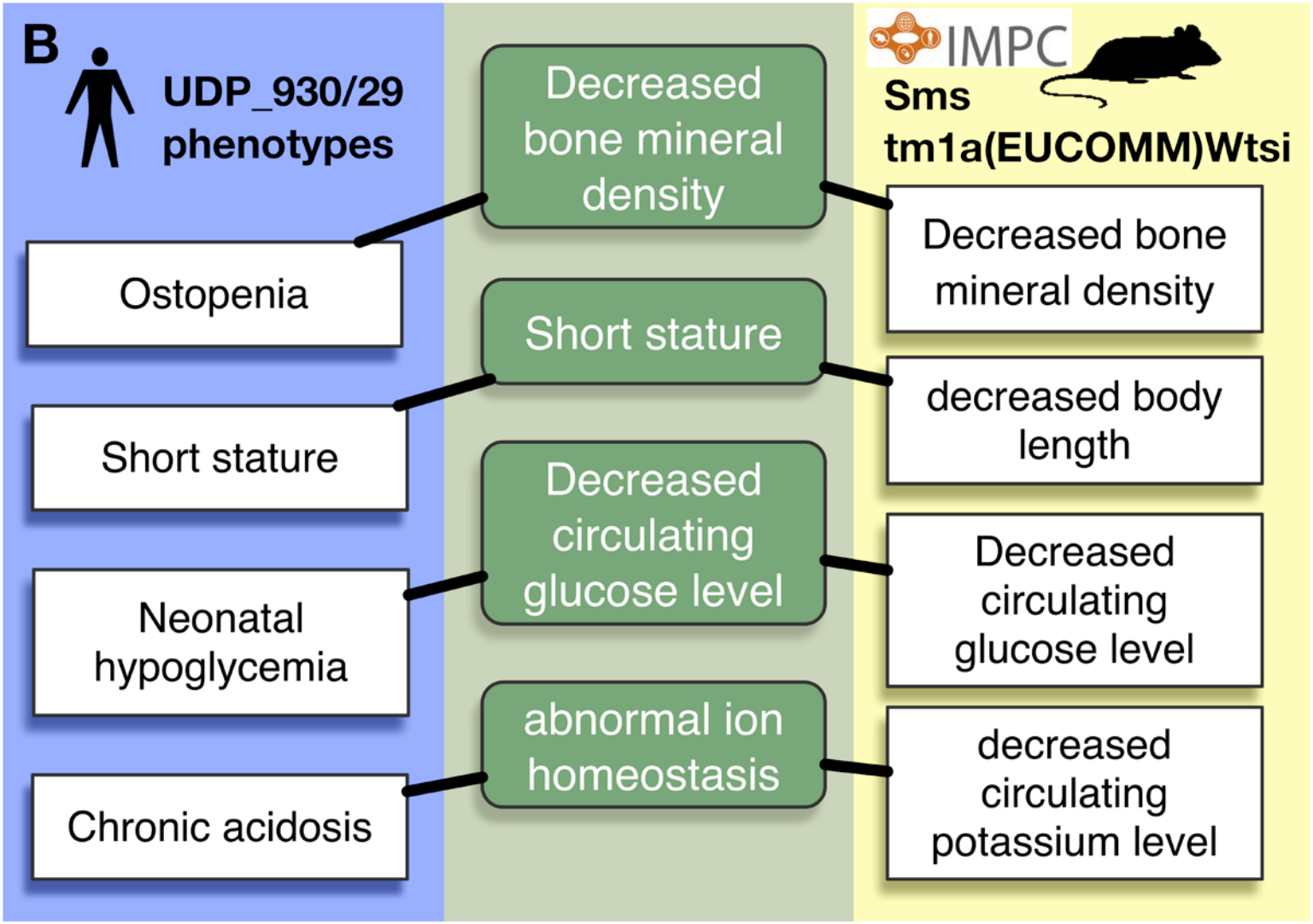
=> Use of genotype, phenotype and inheritance data together provide best prioritization

Integration into UDP pipeline

- 4/23 previously problematic cases received a diagnosis
- One novel disease-gene discovery: York Platelet syndrome and *STIM1*
- Strong candidates identified for other cases: functional validation through mouse and zebrafish modelling
- Several hundred further cases now being analysed

Bone W et al. Genetics in Medicine 2015 (In Press)

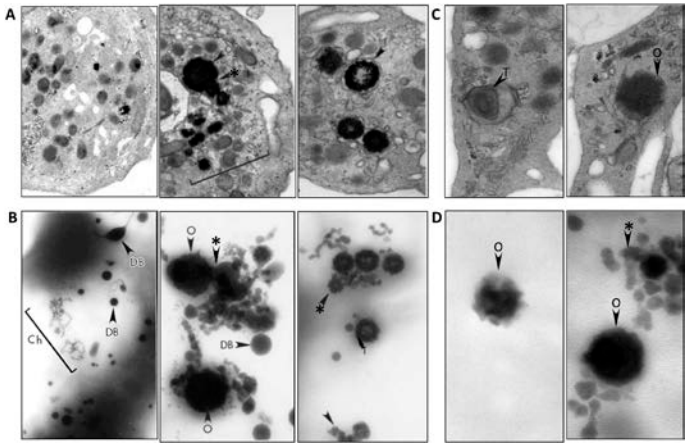
UDP930/929 diagnosed with a *SMS* mutation



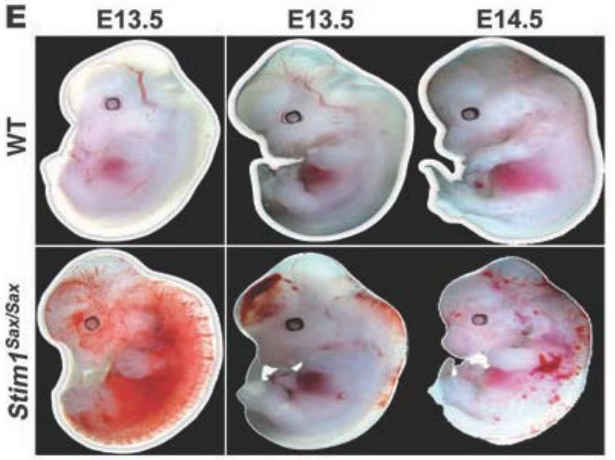
York platelet syndrome and *STIM1*

NIH Undiagnosed Diseases Program

MGI



Markello T et al. Molecular Genetics and Metabolism 2015, 114: 474

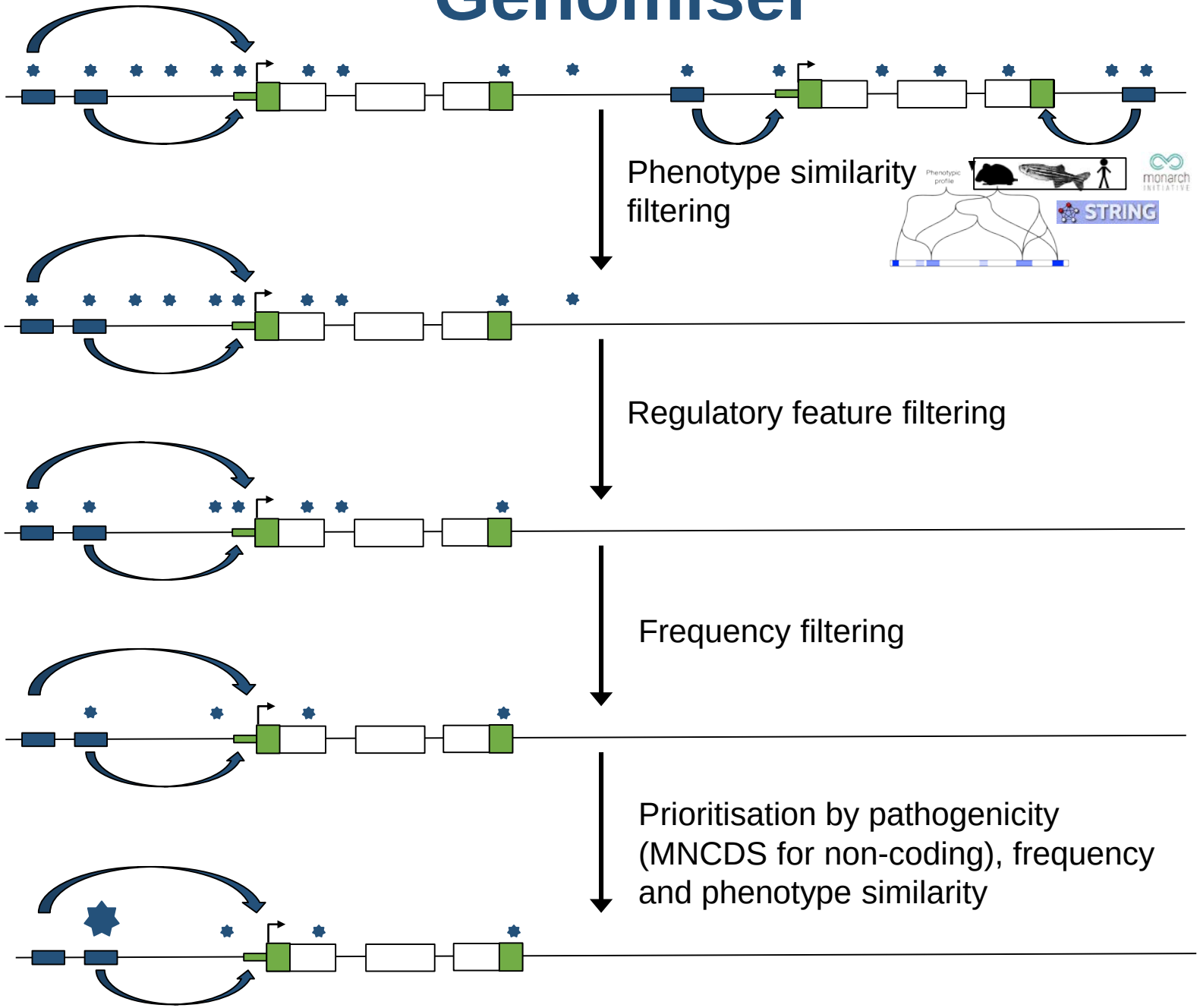


Grosse J, J Clin Invest 2007 117: 3540-50

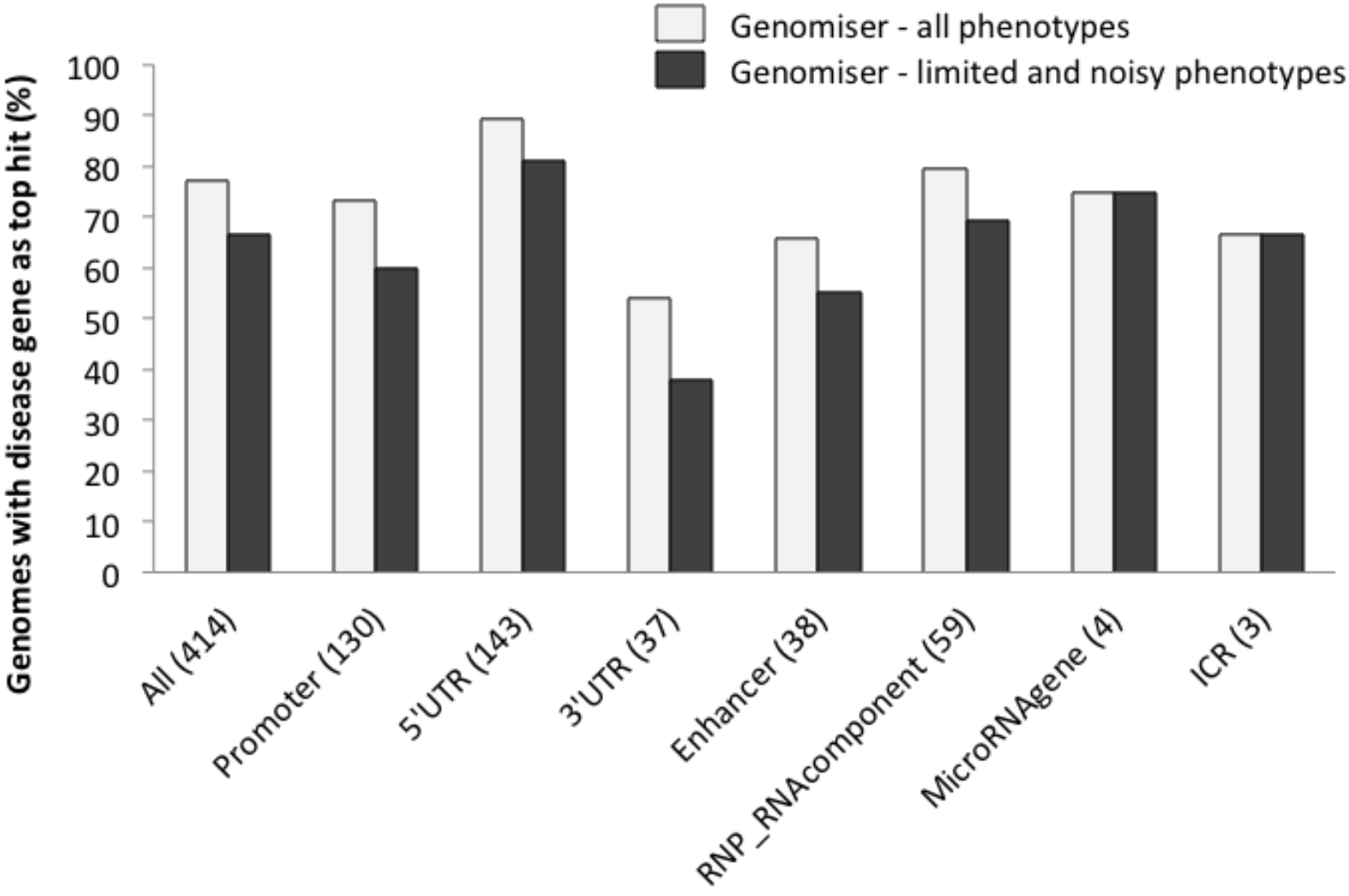
Impaired platelet aggregation (HP:0003540) — Abnormal platelet activation (MP:0006298)

Thromocytopenia (HP:0001873) — Thrombocytopenia (MP:0003179)

Genomiser


Phenotype similarity filtering
Phenotypic profile
monarch INITIATIVE
STRING
Regulatory feature filtering
Frequency filtering
Prioritisation by pathogenicity (MNCDS for non-coding), frequency and phenotype similarity

Mendelian regulatory mutations



Exomiser software suite

- **How-To guide**: **Nature Protocols** 2015 (In Press)

- **PHIVE**: Robinson PN et al. **Genome Research** 2014
- **hiPHIVE:** Bone W et al. **Genetics in Medicine** 2015 (In Press)
- **PhenIX**: Zemojtel T et al. **Science Translational Medicine** 2014
- **ExomeWalker**: Smedley D et al. **Bioinformatics** 2014

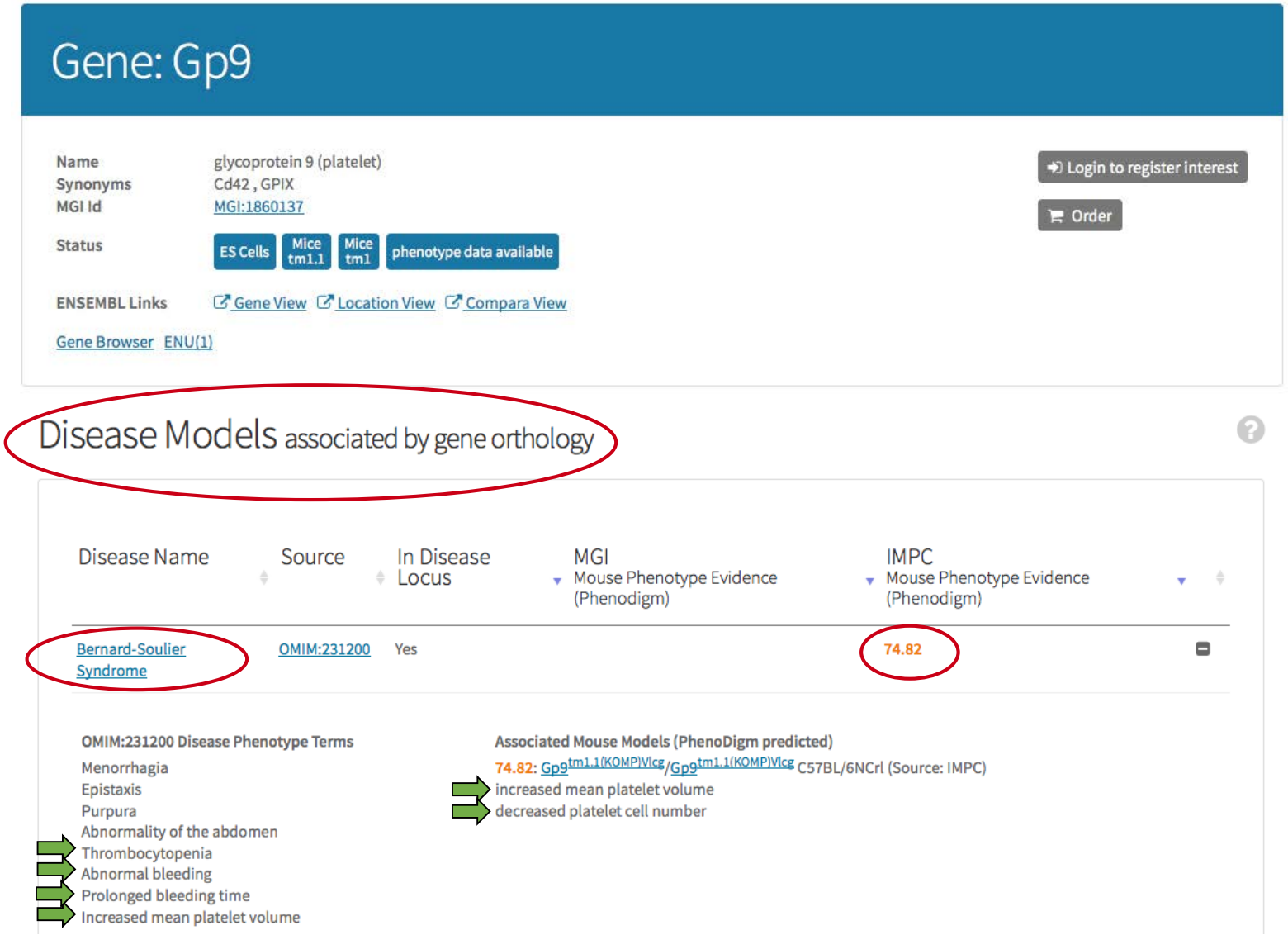
Models for functional validation: NIH KOMP2 and IMPC

- 530 genes associated with a Mendelian disease now have a phenotyped IMPC line

- Potential **new disease models** for 85% as never had a mouse disease model described in literature and 24 already showing phenotype similarity from partial results on the IMPC broad screen

- 75 **novel disease gene candidates** from phenotypic similarity where human ortholog lies in correct linkage locus

First Bernard-Soulier mouse model



Gene: Gp9

Name: glycoprotein 9 (platelet)
Synonyms: Cd42 , GPIX
MGI Id: MGI:1860137

Status: ES Cells | Mice tm1.1 | Mice tm1 | phenotype data available

ENSEMBL Links: Gene View | Location View | Compara View

Gene Browser ENU(1)

Login to register interest

Order

Disease Models associated by gene orthology

Disease Name	Source	In Disease Locus	MGI Mouse Phenotype Evidence (Phenodigm)	IMPC Mouse Phenotype Evidence (Phenodigm)
Bernard-Soulier Syndrome	OMIM:231200	Yes		74.82

OMIM:231200 Disease Phenotype Terms

- Menorrhagia
- Epistaxis
- Purpura
- Abnormality of the abdomen
- Thrombocytopenia
- Abnormal bleeding
- Prolonged bleeding time
- Increased mean platelet volume

Associated Mouse Models (PhenoDigm predicted)

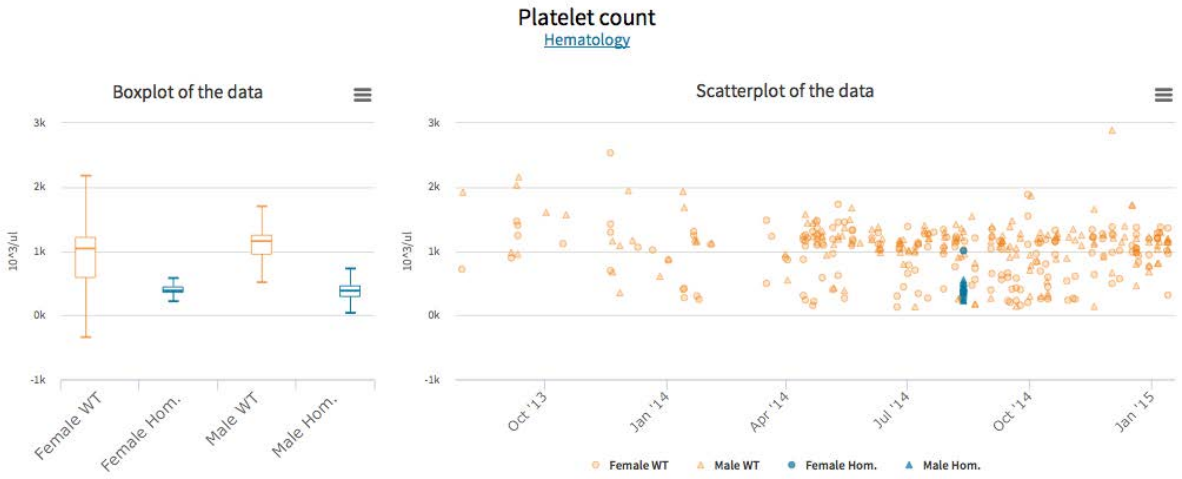
74.82: $Gp9^{tm1.1(KOMP)Vlcg}/Gp9^{tm1.1(KOMP)Vlcg}$ C57BL/6NCrl (Source: IMPC)

- increased mean platelet volume
- decreased platelet cell number

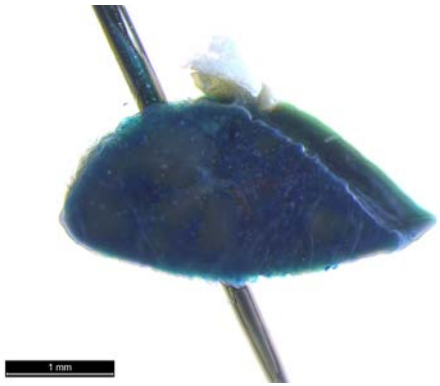
#231200

BERNARD-SOULIER SYNDROME; BSS

CATEGORY	SUBCATEGORY	FEATURES
Inheritance	-	Autosomal recessive
Head and Neck	Nose	Epistaxis
Abdomen	Gastrointestinal	Hemorrhage
Genitourinary	Internal Genitalia (Female)	Menorrhagia
Skin, Nails, Hair	Skin	Purpura
Hematology	-	Congenital bleeding diathesis Large platelets Mild thrombocytopenia
Laboratory Abnormalities	-	Prolonged bleeding time Reduced platelet glycoprotein Ib complex Normal platelet aggregation with ADP, collagen, epinephrine Absent platelet agglutination in presence of ristocetin



Spleen lacZ



Thrombocytopenia (MP:0003179)

First bone mineral QTL18 mouse model

Disease: Bone Mineral Density Quantitative Trait Locus 18

Name	Bone Mineral Density Quantitative Trait Locus 18
Synonyms	OSTEOPOROSIS AND OSTEOPOROTIC FRACTURES, SUSCEPTIBILITY TO
Locus	Xq23
Associated Human Genes	PLS3
Mouse Orthologs	Pls3
Source	OMIM:300910

Mouse Models associated by gene orthology

Mouse Gene Symbol	Disease Gene Ortholog	MGI Phenotype Similarity Score	IMPC Phenotype Similarity Score
Pls3	PLS3		68.88

OMIM:300910 Disease Phenotype Terms

Osteopenia
Osteoporosis

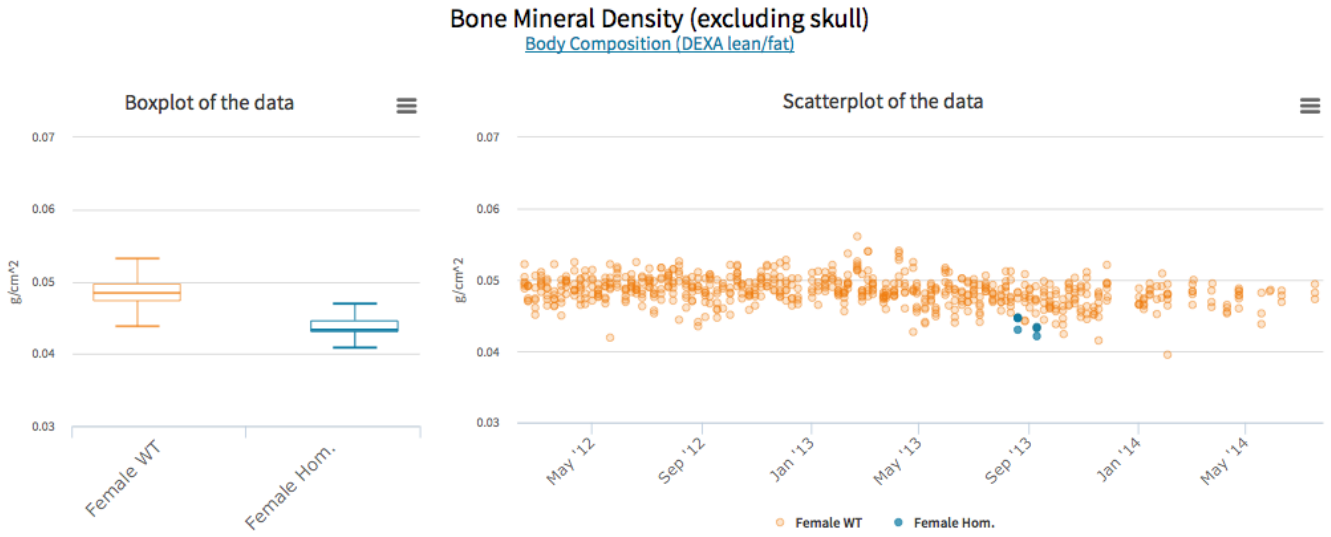
Associated Mouse Models (PhenoDigm predicted)

68.88: / C57BL/6NTac (Source: 3i,IMPC)
decreased bone mineral density
decreased monocyte cell number
increased circulating alkaline phosphatase level

First bone mineral QTL18 mouse model

Disease: Bone Mineral Density Quantitative Trait Locus 18

Name	Bone Mineral Density Quantitative Trait Locus 18
Synonyms	OSTEOPOROSIS AND OSTEOPOROTIC FRACTURES, SUSCEPTIBILITY TO
Locus	Xq23
Associated Human Genes	PLS3
Mouse Orthologs	Pls3
Source	OMIM:300910



Decreased bone mineral density (MP:0000063)

Novel candidate for isolated microphthalmia, with cataract, 1

Disease: Microphthalmia, Isolated, With Cataract 1

Name	Microphthalmia, Isolated, With Cataract 1
Synonyms	CATARACT, CONGENITAL, WITH MICROPHTHALMIA; CATM
Locus	16p13.3
Associated Human Genes	-
Mouse Orthologs	-
Source	OMIM:156850

Mouse Models associated by gene orthology

No mouse models associated with OMIM:156850 by orthology to a human gene.

Potential Mouse Models predicted by phenotypic similarity

Show 10 entries
Search:

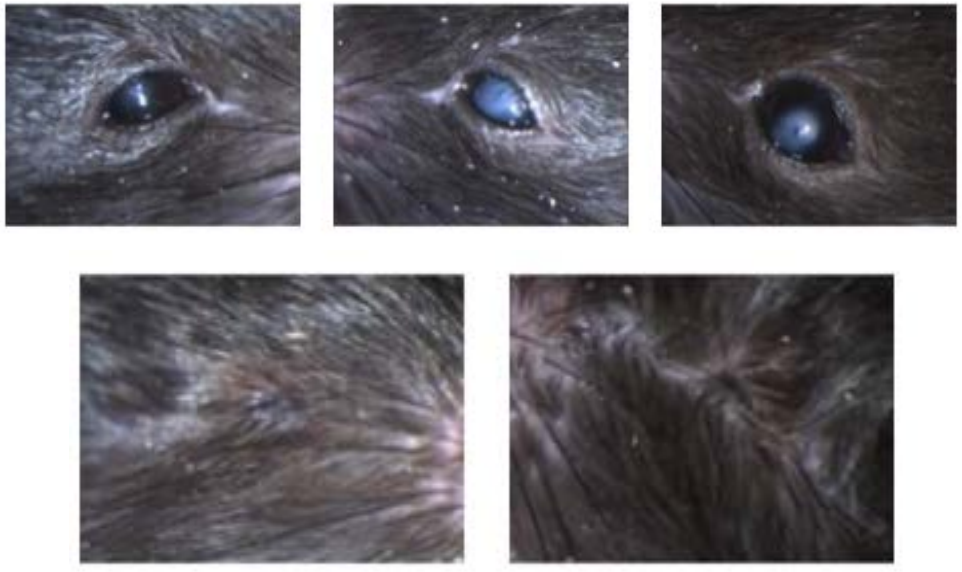
Mouse Gene Symbol	In Disease Locus	MGI Phenotype Similarity Score	IMPC Phenotype Similarity Score	
Slx4	Yes	53.97	69.5	⊕

Novel candidate for isolated microphthalmia, with cataract, 1

%156850

MICROPHTHALMIA, ISOLATED, WITH CATARACT 1; MCOPCT1

CATEGORY	FEATURES
Eyes	Cataract
	Microphthalmia
	Nystagmus
	Miosis
	Strabismus



Associated Mouse Models (PhenoDigm predicted)

69.5: / B6J.129S2.B6N (Source: EuroPhenome,MGP)

abnormal cranium morphology
decreased body weight
anophthalmia
abnormal cornea morphology
corneal opacity
abnormal eye morphology
abnormal eye size
decreased platelet cell number
abnormal placement of pupils

Conclusions

- Semantic phenotype comparisons greatly improve diagnosis and candidate gene identification as well as highlighting good disease models
- Inclusion of mouse and fish phenotypes along with guilt by association from PPA data is critical, especially for novel disease gene discovery
- Our results clearly show the value of collecting deep clinical phenotype data for translational bioinformatics

Future challenges

- Inclusion of phenotype frequency data
- Inclusion of negative phenotype data
- Certain phenotypes, e.g. behavioral, are not well covered by mouse/fish and/or our algorithms => incorporate new ontological approaches and species e.g. primates
- Common disease

Acknowledgments

NIH-UDP
- **William Bone**
- Murat Sincan
- David Adams
- Amanda Links
- Joie Davis
- Neal Boerkoel
- Cyndi Tifft
- Bill Gahl

OHSU
- **Melissa Haendel**
- Nicole Vasilesky
- Matt Brush
- Bryan Laraway
- Shahim Essaid
- Kent Shefchek

EBI
- **Terry Meehan**
- Helen Parkinson
- Jeremy Mason
- Chao-Kung Chen

Lawrence Berkeley
- **Nicole Washington**
- Suzanna Lewis
- **Chris Mungall**

MRC Harwell
- **Ann-Marie Mallon**
- Hugh Morgan
- Ahmad Retha
- Andrew Blake
- Armida Di Fenza
- Duncan Sneddon
- Gagarine Yaikhom
- Henrik Westerberg
- Luis Santos
- Tanja Fiegel

Sanger
- Bill Skarnes
- Anika Oelllrich
- **Jules Jacobson**
- Irene Papatheodorou

Toronto
- Marta Girdea
- Sergiu Dumitriu
- Heather Trang
- Orion Buske
- Mike Brudno

JAX
- Cynthia Smith

Charité
- Sebastian Kohler
- Sandra Doelken
- Max Schubach
- **Peter Robinson**

NIH: U54 HG006370-01
NIH Office of Director: 1R24OD011883
NIH-UDP: HHSN268201300036C, HHSN268201400093P

monarch INITIATIVE

IMPC MPI2