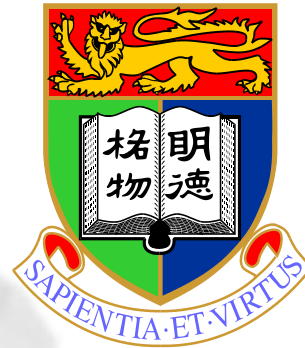


Uncovering the genetic lesions underlying the most severe form of Hirschsprung (HSCR) disease by whole genome sequencing (WGS): a pilot study in 8 family trios



Mercè GARCIA-BARCELO

Department of Surgery

The University of Hong Kong

16th June 2017

What is Hirschsprung disease (HSCR)?:

- Congenital disorder ▶ abnormal development of the gut
- Lack of ganglion cells in the distal gut ▶ no peristalsis
- Surgery only available treatment ▶ lethal if untreated
- Heterogeneous phenotype
- Boys 4 times more affected than males



Aganglionic segment

0 SHARE



NOW READING

Doctors remove 13kg of faeces from constipated man in China



Doctors remove 13kg of faeces from constipated man in China

22-year-old diagnosed with Hirschsprung's disease said he'd been suffering from constipation, stomach pain all his life

PUBLISHED : Tuesday, 13 June, 2017, 2:55pm

UPDATED : Tuesday, 13 June, 2017, 2:55pm



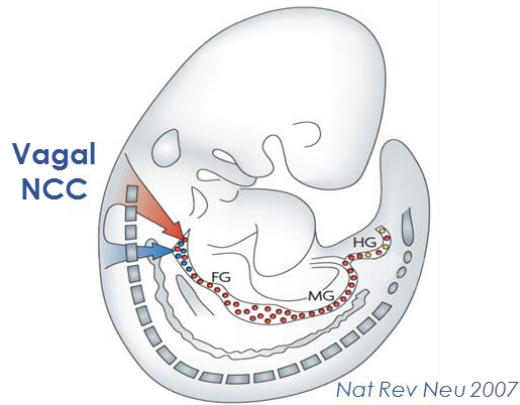
Caucasians 1.5/10,000

Asians 2.8/10,000



How/why does it happen?

Neural crest cells (NCC) give rise to enteric neurons (EN)

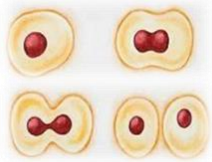


Perturbation



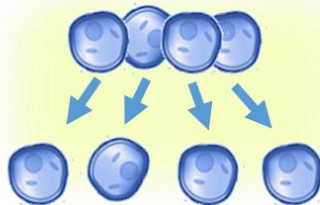
HSCR

Proliferation



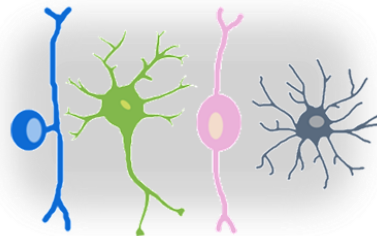
Correct population size of stem cells

Migration



Complete colonization

Differentiation

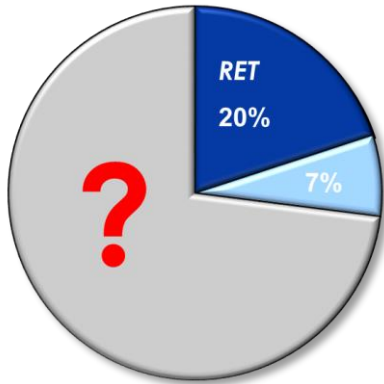


Full diversity of neurons & glia

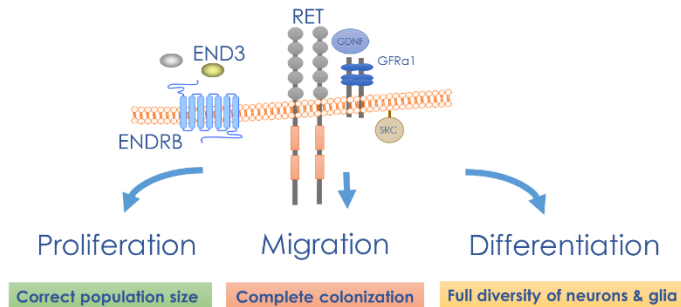
DNA variants impairing genes involved in EN development

What do we know about the genetic risk factors?

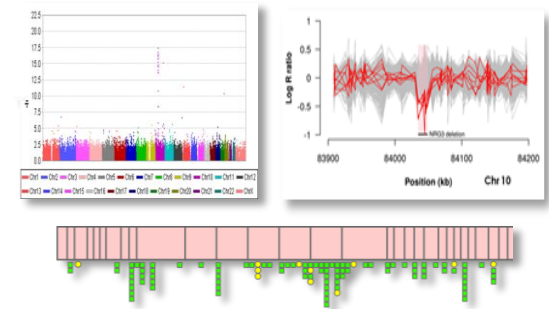
RET main gene



Other genes of interacting pathways involved in ENS



Common and rare variants involved



Oligogenic

Genetically heterogeneous

Heterogeneous phenotype

- Long segment (20%)
- Short segment (80%)

Variable presentation

- Familial (20%)
- Sporadic (80%)

Various disease mechanisms

Affecting either:

- Proliferation
- Migration
- Differentiation

What else do we know about the genetic risk factors?

From our and others' observations:



Variants:

- Any type (small/large variants)
- Anywhere in the genome ► relevance of regulatory processes during development

Not all patients are accounted for by the known genes



There exist severely affected HSCR patients, sporadic, that cannot be accounted by rare variants in known genes ► suitable for search of new genetic factors

- HSCR patients affected with the **most severe phenotype**
- Born to unaffected patients ► **sporadic**
- **Devoid** of damaging variants in the coding sequences of known HSCR genes
- Whole genome sequencing
- Trio-based approach ► **detection of *de novo* variants**



Trio 1



Trio 2



Trio 3

.....



Trio 9

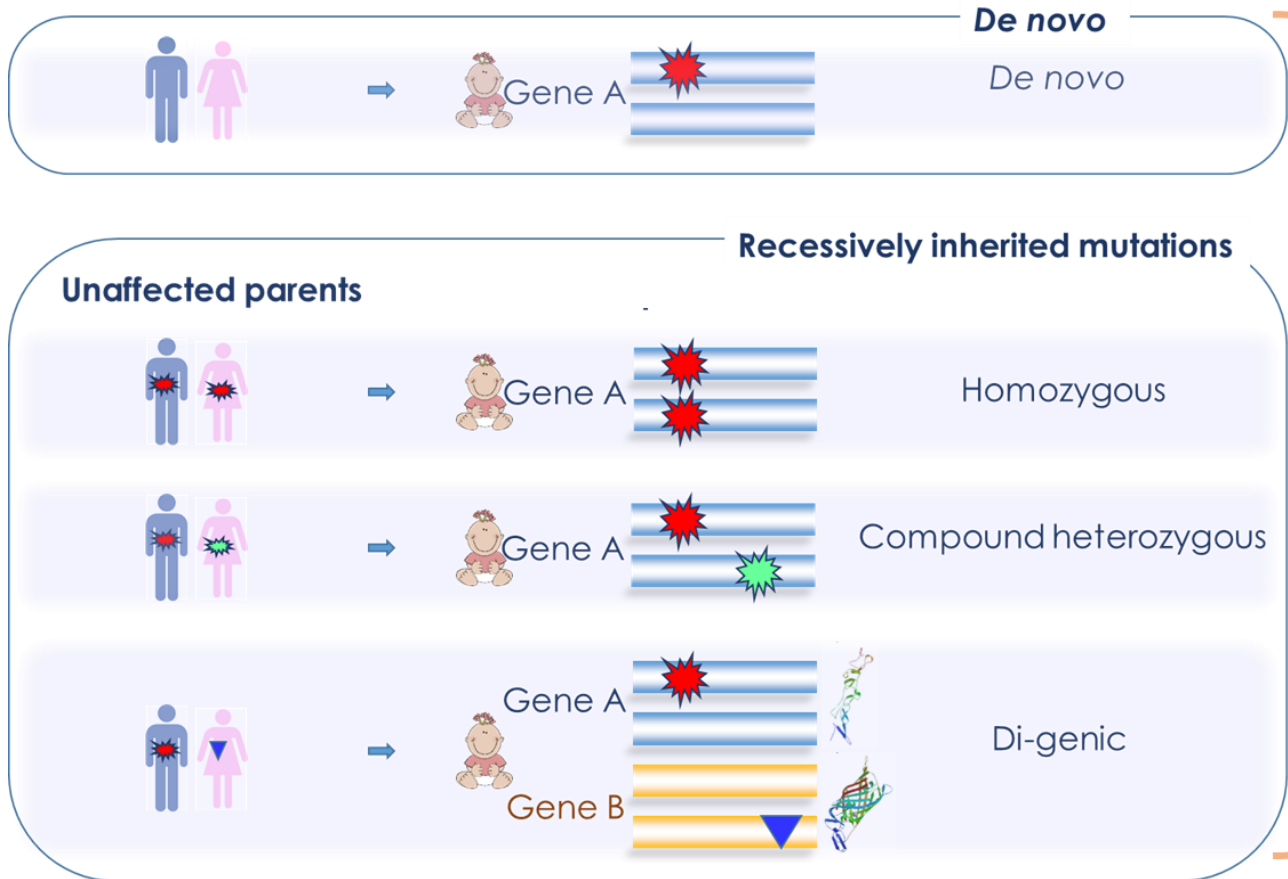
Why whole genome sequencing?

To cover in both coding and regulatory regions:

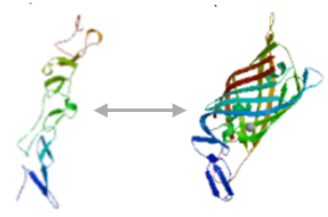
- Single nucleotide variants (SNV)
- Small deletions and insertions (Indels)
- Rare structural variations/Copy number variants (CNV)

Platform Variants	Common variants array	Whole exome sequencing (WES)	Whole genome sequencing (WGS)
Common variants	✓	✓ (only coding)	✓
Rare variants		✓ (only coding)	✓
CNV	✓		✓

Long segment HSCR = Rare + Sporadic disorder



Rare/absent in the population



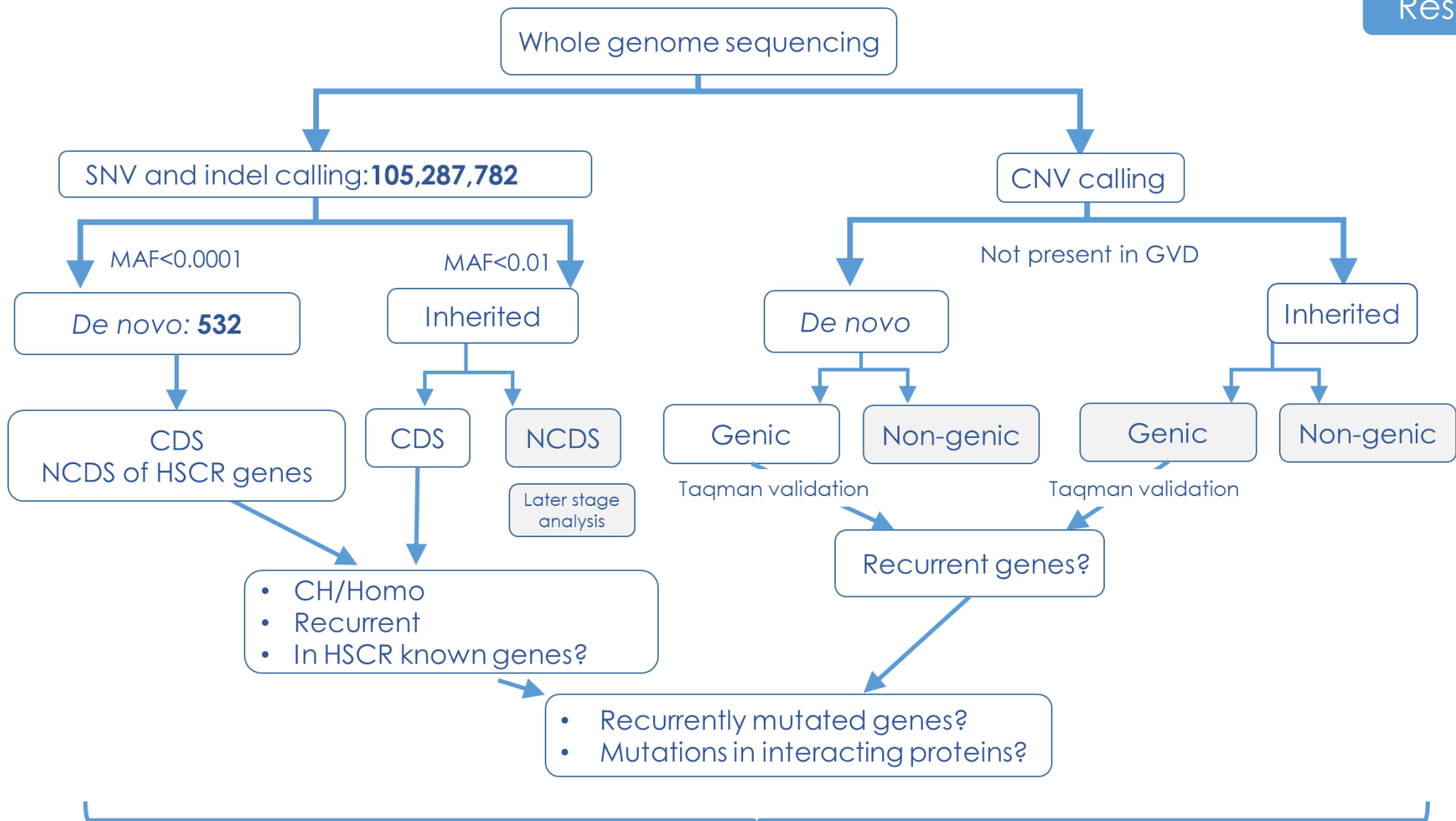
MAF (Minor Allele Frequency)

De novo MAF < 0.0001 % or Novel

Recessive: < 0.01% or Novel

Compound heterozygous: < 0.011% or Novel

Digenic : < 1% or Novel

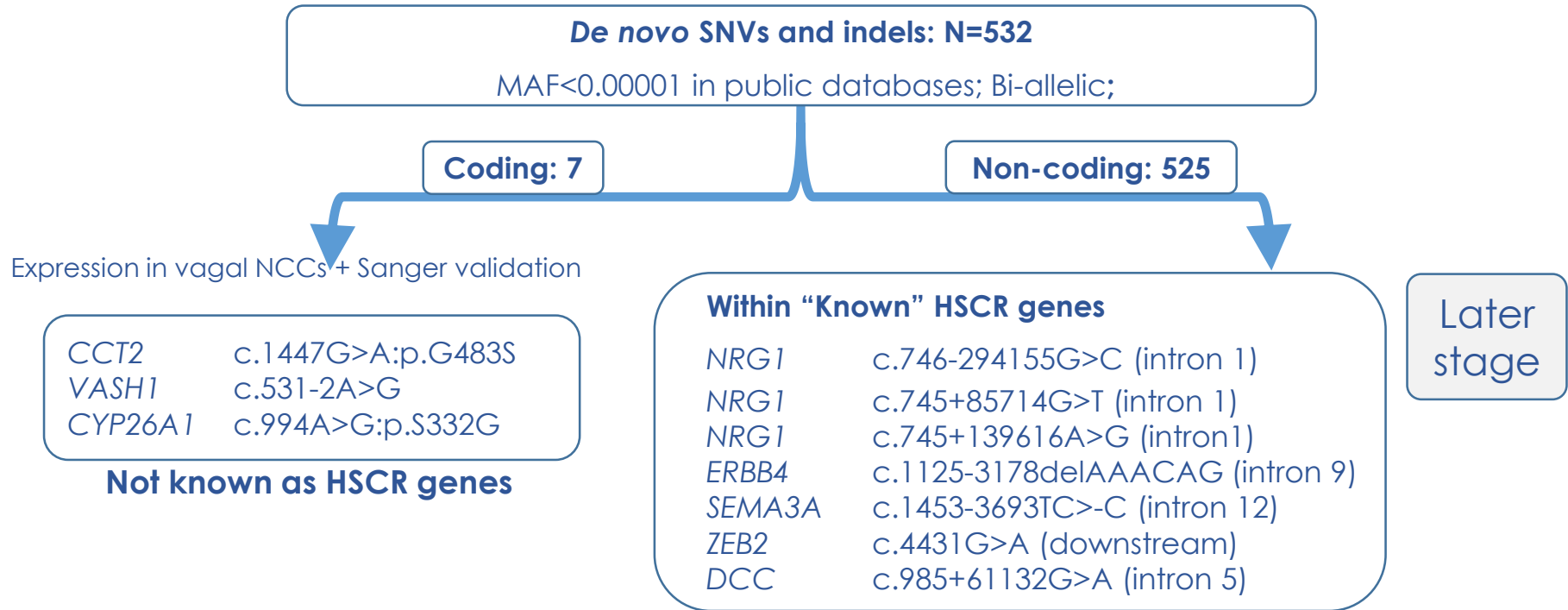


SNV: Single nucleotide variant
 Indel: Insertions and deletions < 100bp
 CH: compound heterozygous
 CDS: Coding sequence
 NCDS: Non-coding sequence
 CNV: copy number variant
 GVD: Genome Variation Database

Patients mutational profile

Shared genetic/pathways features

Filtering and selection of *de novo* single nucleotide and indel variants



Amongst the 10-15% genes most intolerant to changes human genes

- *CCT2*: formation of the primary cilia. Bardet-Biedl syndrome –can co-exist with HSCR-
- *VASH1*: secreted protein angiogenesis regulation.
- *CYP26A1*: retinoic acid (RA)-metabolizing enzyme.

Exonic homozygous mutations: **none in “HSCR genes”**

Patients	Genes	Variant	Gene function
HD09C	<i>FOCAD</i>	p.A1709T	Potential tumor suppressor in gliomas
HK164C	<i>BRD1</i>	p.A689V	Histones H3 and H4 acetylation
	<i>C5orf42</i>	p.G2168D	Transmembrane protein
	<i>GINS4</i>	p.V171M	Initiation of DNA replication
	<i>GLRX3</i>	p.I330V (cu)	Crucial regulator of cellular iron homeostasis
	<i>HK2</i>	p.A352T	Glycolysis, gluconeogenesis
	<i>ITGB5</i>	p.L94V	Integrin. Extracellular matrix (ECM)
	<i>NEK1</i>	p.R586H	Centrosomal complex. Microtubule assembly
	<i>PLAT</i>	p.Y425H	Direct role in neuronal migration
	<i>PPP2R3A</i>	p.Y431C	Intracellular signaling
	<i>RRP7A</i>	p.T81	Ribosomal RNA Processing
	<i>STXBP5L</i>	p.R618Q	GTPase activator activity
HK180C	<i>USP42</i>	p.D1220G (cu)	Deubiquitinating enzyme
	<i>VRK2</i>	p.R491H	Effector of signaling pathways that regulate apoptosis/tumor cell growth
HK180C	<i>XRN2</i>	p.A529T	5'-3' Exoribonuclease 2
HK96C	<i>BICD2</i>	p.R398W	Essential for motor neuron physiology
VH105C	<i>PLEKHA4</i>	p.R571H	Binds to phosphatidylinositol 3-phosphate
VH108C	<i>SEMA7A</i>	p.V320I	Integrin-mediated signaling. Focal adhesion. Neuronal functions

cu: case unique variant

***PLAT*: direct role in facilitating neuronal migration**

Compound heterozygous variants (CH)

RADIL: knockdown of *radil* in zebrafish results in multiple defects in pigment cells and enteric neurons

Patients	Gene	Variants (paternal/maternal)	Disorders in humans
HD09C	BRD4	c.-7165+1G>A/p.E1326D	Translocation breakpoints in 2 patients with carcinomas
	MAGI3	p.D1059G 9(cu)/p.Y285H	ND
	FGFRL1	p.R241Q/p.A288T	Wolf-Hirschhorn syndrome (WHS)
	SSH2	p.G1378D/p.P1338S(cu)	ND
HK164C	TFR2	p.A75V/p.S506G	Hemochromatosis, Type 3
	DOCK8	p.C871W+p.Y2001* (cu)/p.S1177L	Non-syndromic intellectual disability
	CDC14A	p.R582H/p.P125H	Deafness, Autosomal Recessive
	FRAS1	p.N3119S/p.K1873Q	Cryptophthalmos, cutaneous syndactyly and genitourinary anomalies. AR
HK180C	SLC24A1	p.E697G/p.S521N	Congenital stationary night blindness (AR)
	CUL7	p.L1710V/p.R707C	3M syndrome (AR). Dubowitz's syndrome
	ACOX2	p.A504T/p.R409H	ND
	ARFGEF3	p.D87Y (cu)/p.E902Q (cu)	ND
HK97C	NACAD	p.Q1391E/p.G407R (cu)	ND
	PCNT	p.A1495V / p.R1821W	Microcephalic osteodysplastic primordial dwarfism type 2. Primary autosomal recessive microcephaly type 6. Seckel syndrome type 4
HK9C	LAMA5	p.E2378K/p.E2665K	ND
	CMYA5	p.S308F/p.G450D	Cardiomyopathy
	MGAM	p.M1073V/p.D1454G	Intestinal disaccharidase deficiency.
	RADIL	p.V337M/p.L200P (cu)	ND
VH105C	CUL7	p.S999C (cu)/p.V1691I	3M syndrome (AR). Dubowitz's syndrome
	ZSWIM4	p.R185W (cu)/p.Q267H	ND
VH106C	SYNE1	p.S2126F/p.R5617*	SYNE1-Related Autosomal Recessive Cerebellar Ataxia. Emery-Dreifuss muscular dystrophy 4, autosomal dominant
VH108C	RFC2	p.E207K/p.P22S	Gene in the Williams-Beuren Syndrome critical region. Infantile hypercalcemia. Aortic stenosis syndrome

Genes recurrently mutated

- Recurrently mutated genes?
- Mutations in interacting proteins?

Patients mutational profile

Shared genetic/pathways features

Considered genes with rare SNVs and/or CNVs

De novo
Homozygous
Compound heterozygous

Identified in this study
-seed genes-

HSCR / ENS related genes (117 genes)

21 genes have different rare variants **in more than one patient**

Functional overlap among genes

Di/oligo-genic model where variants in two or more interacting genes co-exist in a patient

- Recurrently mutated genes?
- Mutations in interacting proteins?

Patients mutational profile

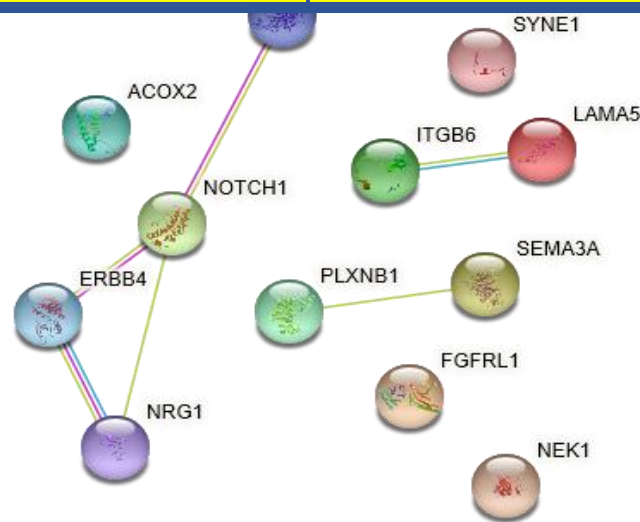
Shared genetic/pathways features

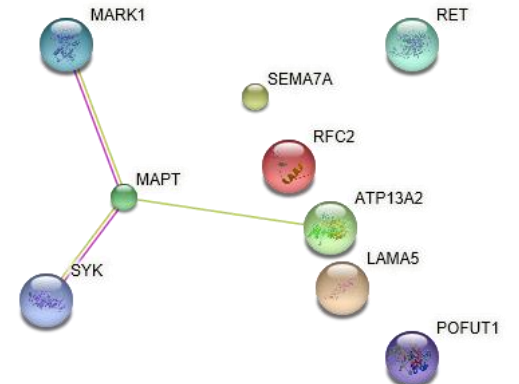
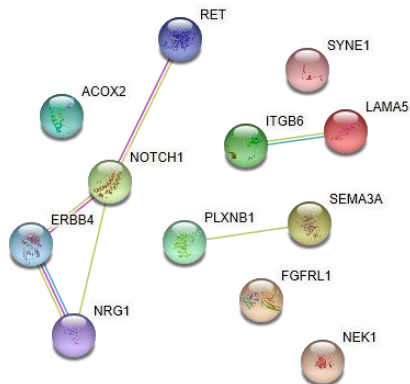
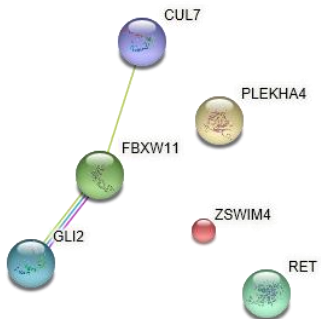
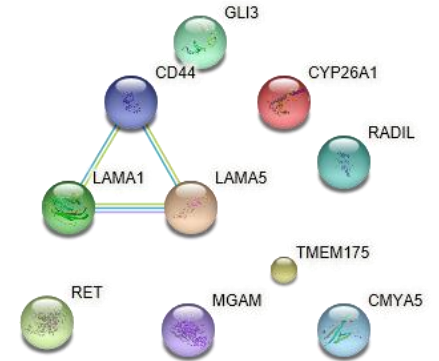
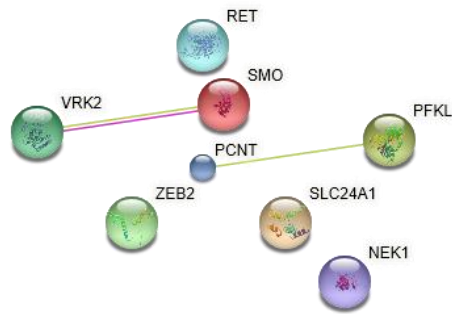
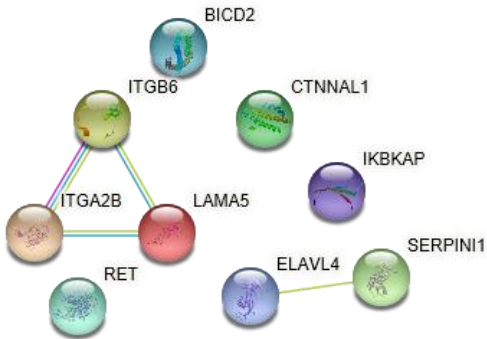
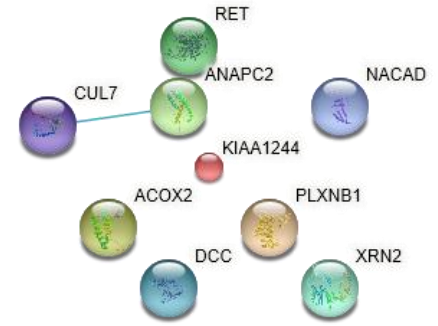
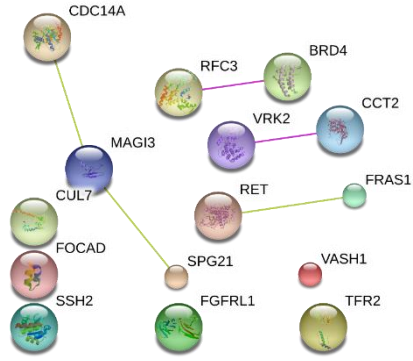
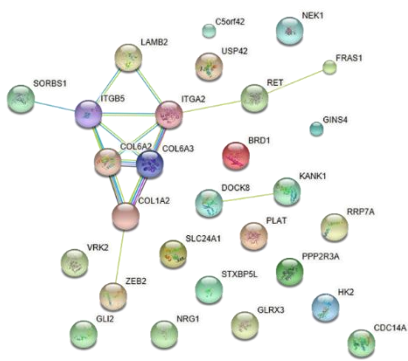
Included genes encoding **interacting partners** to **any of the seed genes** / ENS genes



Genetic profile of each patient

Patients	Gene	Variant	Reason for gene selection	Type
VH106C	SYNE1	p.S2126F/p.R5617*	Seed	CH
	ERBB4	c.1125-3178delAAACAG (intron 9)	Seed and ENS-gene	NCDS <i>de novo</i>
	NRG1	c.746-294155G>C (intron 1)	Seed and ENS-gene	NCDS <i>de novo</i>
	NOTCH1	p.R1330H (cu)	ENS-gene	Paternaly inherited
	ACOX2	p.R88Q	Seed -CH in patient HK180-	Maternaly inherited
	FGFRL1	p.V98L	Seed -CH in patient HD9-	Paternaly inherited
	NEK1	p.R355G	Seed -Homo in patient HK164-	Paternaly inherited
	LAMA5	p.L2185F	Seed - CH in patient HK9C-	Paternaly inherited
	ITGB6	p.R499H	Interacting partner of LAMA5	Maternaly inherited
	PLXNB1	p.L1686M	ENS-gene	Maternaly inherited
SEMA3A	c.1453-3693TC>-C (intron 12)	Seed and ENS-gene	NCDS <i>de novo</i>	





Is there anything in common? Any shared genetic component?

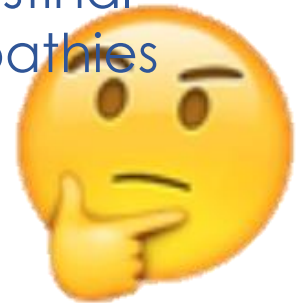
Each patient has variants in at least two interacting and biologically plausible genes

Extra Cellular Matrix-receptor interaction pathway (ECM-receptor) has significantly more interactions than expected ($p=1.5 \times 10^{-11}$)

- *Could any EMC molecule be used as a target?*

Detected variants in **schizophrenia, autism** (*NRG1, NRG3, ERBB4, SEMA3A, PLXNB1, DOCK8, CALN1, NBPF*) and **ciliopathies** (*CCT2*) genes

- Frequently observed association between intestinal dysmotility and psychiatric disorders and ciliopathies (Bardet-Biedl; Joubert syndromes)



Pathological alterations affecting pathway(s) **shared by more than one disorder may underlay apparently unrelated diseases**

- How can we make use of this?
- Predictable phenotypic pattern?

Pilot study

First WGS in HSCR

Very instrumental in setting a pipeline at HKU



Human Medical Research Fund 01121516

The University of Hong Kong

Department of Surgery

Clara Sze-man Tang
Sunny Xuehan Zhuang
Michelle Yu
Man-ting So
Paul Tam



Clara Tang

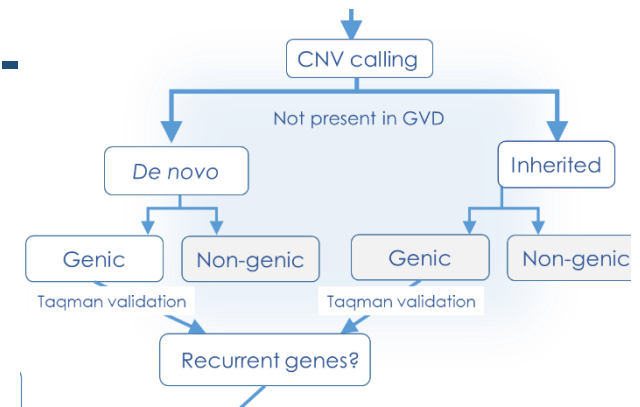
Centre for Genomic Sciences

Pak Sham
Stacey Cherny



Sunny Zhuang

Preliminary CNV data –genetic profile-



	De novo		Inherited	
	Genes deleted	Genes duplicated	Genes deleted	Genes duplicated
HD09C		CALN1 (6K; intronic) <i>No in 1K project</i> CALN1 (29K; intronic)	GFRA2 (36K; exonic)	NOTCH2 (5K; exonic)
HK164C	NBPF6(3K, exonic)	NBPF8, NBPF9 (9K; exonic)	GFRA2 (36K; exonic)	
Recurrent gene. Duplicated in another HSCR patient	PPM1L (3.5Kb, intronic) No in 1K project		KIAA1279 (5K; exonic) MGAM (28K; intronic)	
HK180C	NRG3 (2K; intron 1 NM_001010848.3)		GFRA2 (36K; exonic) NRG3 (2K; intron 2; NM_001010848.3) MGAM (28K; intronic)	NOTCH2 (12K; exonic)
HK96C	KLRC3 (6Kb, exonic)		GFRA2 (36K; exonic)	NOTCH2 (12K; exonic)
HK97C	KIR2DL3, KIR3DL3 (38K, exonic) KIR2DL1, KIR2DL4, KIR3DL1 (37K, exonic)			NOTCH2 (12K; exonic)
HK9C	NBPF20 (116K; exonic)		GFRA2 (36K; exonic)	
VH105C	NBPF20 (8K; exonic)		GFRA2 (36K; exonic)	NOTCH2 (12K; exonic)
Recurrent. In different patient	KIR2DL1, KIR2DL4, KIR3DL1 (37K, exonic)			
Recurrent. In different patient. Duplicated.	LILRA6(2Kb, exonic)immunity?		MGAM (28K; intronic)	
VH106C			GFRA2 (36K; exonic) MGAM (28K; intronic)	NOTCH2 (12K; exonic)
VH108C	KLRC3 (4Kb, upstream)		GFRA2 (36K; exonic)	NOTCH2 (12K; exonic)

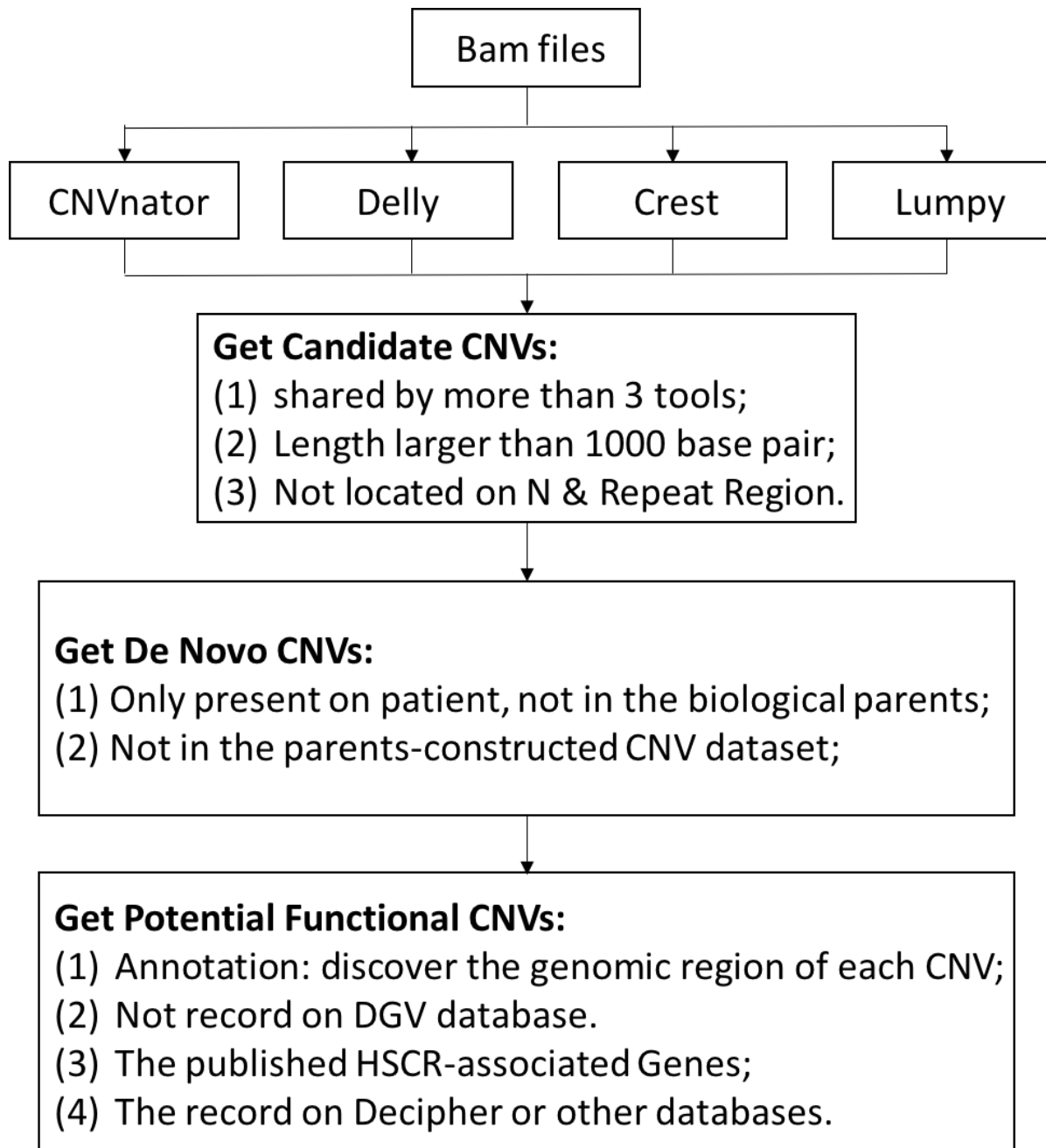
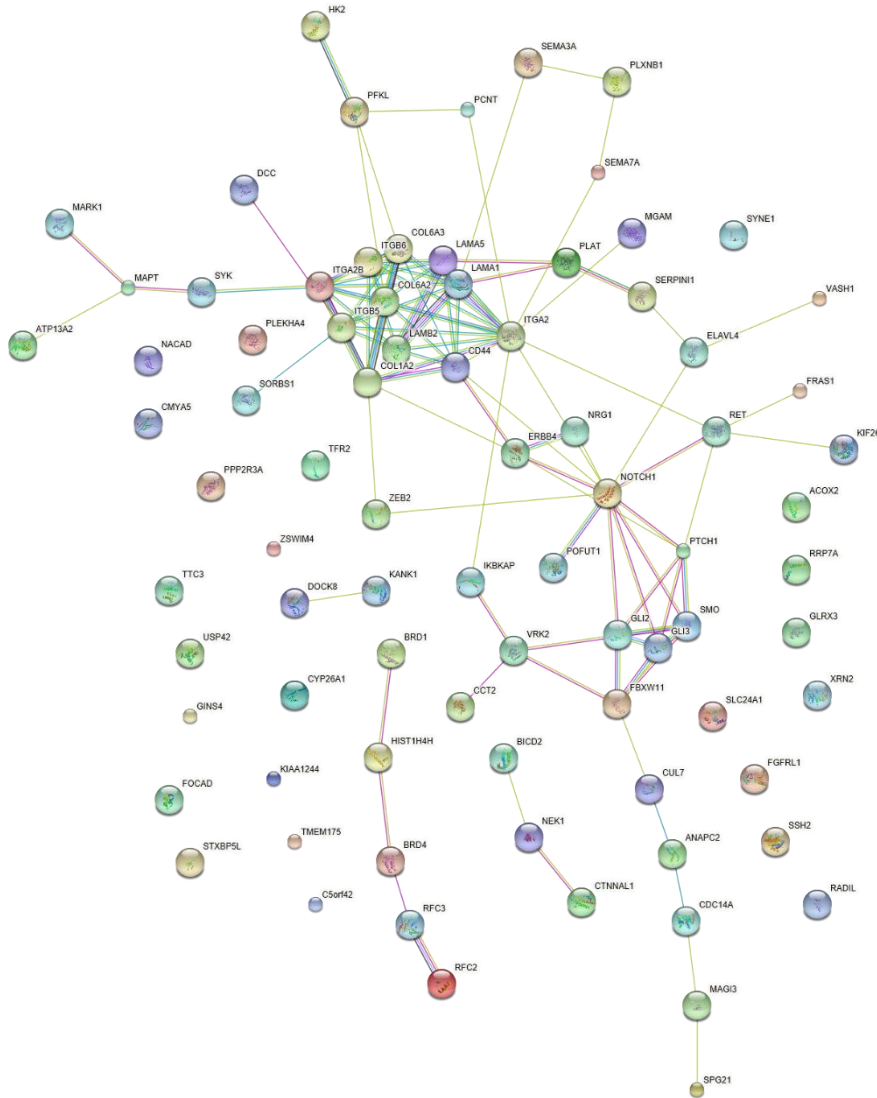


Table S1: Characteristics of the HSCR patients included in the study

Samples	Relationship	Gender	Aganglionosis leng	Associated anomalies/others	Ethnicity
HK9C	Proband	M	Long	Necrotizing enterocolitis; asthma	Chinese
HK9A	Father	M			
HK9B	Mother	F			
HK96C	Proband	M	Total	Parathyroid nodules; bilateral hydrocele; necrotizing enterocolitis	Chinese
HK96A	Father	M			
HK96B	Mother	F			
HK97C	Proband	M	Long	Congenital central hypoventilation syndrome (CCHS); necrotizing enterocolitis ; mild mental retardation; epilepsy	Chinese
HK97A	Father	M			
HK97B	Mother	F			
HK164C	Proband	M	Long	None; consanguineous parents	South-east Asian
HK164A	Father	M			
HK164B	Mother	F			
HD09C	Proband	M	Long	None	Chinese
HD09A	Father	M			
HD09B	Mother	F			
VH105C	Proband	M	Long	None	Vietnamese
VH105A	Father	M			
VH105B	Mother	F			
VH106C	Proband	M	Long	None	Vietnamese
VH106A	Father	M			
VH106B	Mother	F			
VH108C	Proband	M	Long	None	Vietnamese
VH108A	Father	M			
VH108B	Mother	F			
HK180C	Proband	M	Long	None	Chinese
HK180A	Father	M			
HK180B	Mother	F			

All genes



Network Stats

number of nodes: 83
 number of edges: 110
 average node degree: 2.65
 clustering coefficient: 0.647

expected number of edges: 35
 PPI enrichment p-value: 0

your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

Biological Process (GO)

pathway ID	pathway description	count in gene set	false discovery rate
GO:0040011	locomotion	25	8.7e-09
GO:0006928	movement of cell or subcellular component	25	4.3e-08
GO:0009653	anatomical structure morphogenesis	29	5.73e-07
GO:0016477	cell migration	18	1.02e-06
GO:0000902	cell morphogenesis	19	2.73e-06

(more ...)

Cellular Component (GO)

pathway ID	pathway description	count in gene set	false discovery rate
GO:0043234	protein complex	35	0.000142
GO:0043235	receptor complex	10	0.000142
GO:0008305	integrin complex	4	0.00135
GO:0043256	laminin complex	3	0.00135
GO:0044420	extracellular matrix component	6	0.0018

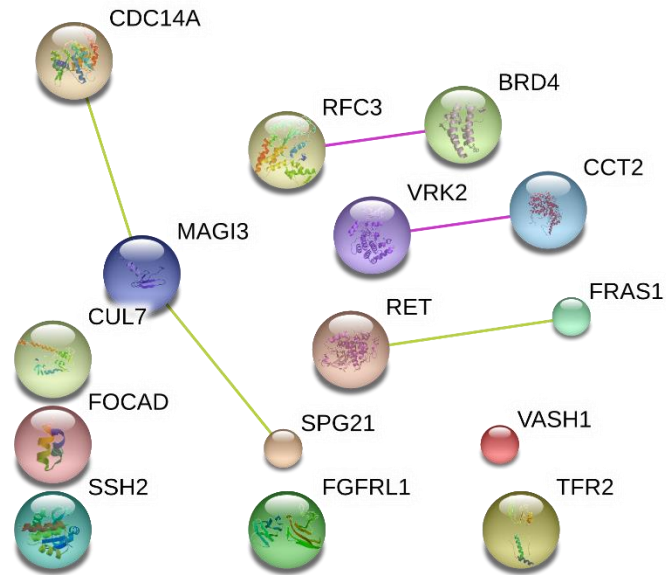
(more ...)

KEGG Pathways

pathway ID	pathway description	count in gene set	false discovery rate
04512	ECM-receptor interaction	11	1.19e-11
04151	PI3K-Akt signaling pathway	12	1.03e-06
04510	Focal adhesion	10	1.03e-06
05200	Pathways in cancer	11	4.92e-06
04340	Hedgehog signaling pathway	5	0.000103

(more ...)

Patient HD9



number of nodes: 15
number of edges: 5
average node degree: 0.667
clustering coefficient: 0.933

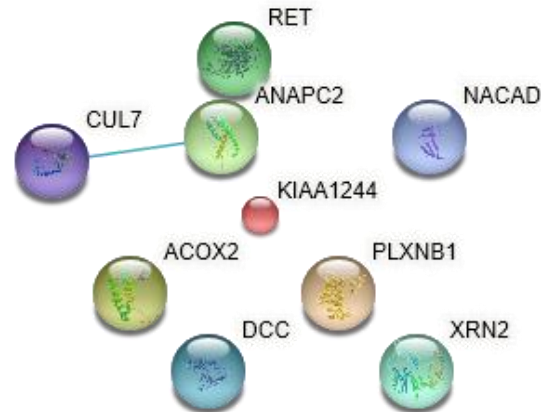
expected number of edges: 1
PPI enrichment p-value: 0.00144

your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

- no significant enrichment detected -

Patient HK180C



Network Stats

number of nodes: 9
 number of edges: 1
 average node degree: 0.222
 clustering coefficient: 1

expected number of edges: 0
 PPI enrichment p-value: 0.286

*your network does **not** have significantly more interactions than expected (what does that mean?)*

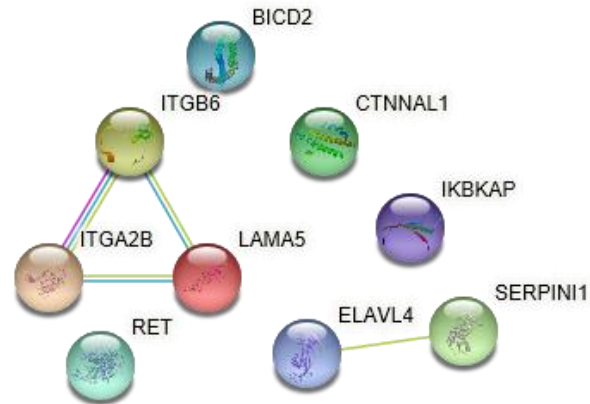
Functional enrichments in your network

Biological Process (GO)

pathway ID	pathway description	count in gene set	false discovery rate
GO:0010769	regulation of cell morphogenesis involved in differentiation	5	0.000693
GO:0010975	regulation of neuron projection development	5	0.000693
GO:0050770	regulation of axonogenesis	4	0.000776
GO:0010976	positive regulation of neuron projection development	4	0.00233
GO:0050773	regulation of dendrite development	3	0.0107
GO:0008361	regulation of cell size	3	0.0218
GO:0022604	regulation of cell morphogenesis	4	0.0239
GO:0010770	positive regulation of cell morphogenesis involved in differentiation	3	0.026
GO:0048799	organ maturation	2	0.0264
GO:0050775	positive regulation of dendrite morphogenesis	2	0.046

(less ...)

Patient HK96C



Network Stats

number of nodes: 9
 number of edges: 4
 average node degree: 0.889
 clustering coefficient: 1

expected number of edges: 0
 PPI enrichment p-value: 0.000412
your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

Cellular Component (GO)

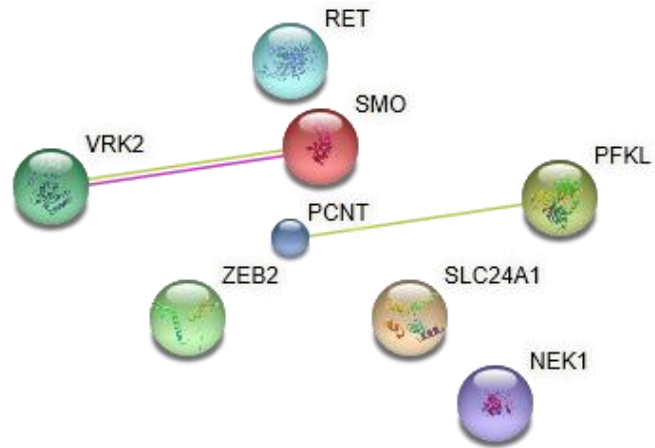
<i>pathway ID</i>	<i>pathway description</i>	<i>count in gene set</i>	<i>false discovery rate</i>
GO:0008305	integrin complex	2	0.0491

KEGG Pathways

<i>pathway ID</i>	<i>pathway description</i>	<i>count in gene set</i>	<i>false discovery rate</i>
04512	ECM-receptor interaction	3	0.00164
04510	Focal adhesion	3	0.0109
04151	PI3K-Akt signaling pathway	3	0.0232
05200	Pathways in cancer	3	0.0232
05222	Small cell lung cancer	2	0.0232
05410	Hypertrophic cardiomyopathy (HCM)	2	0.0232
05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	2	0.0232
05414	Dilated cardiomyopathy	2	0.0232

(less ...)

Patient HK97C



Network Stats

number of nodes: 8
number of edges: 2
average node degree: 0.5
clustering coefficient: 1

expected number of edges: 0

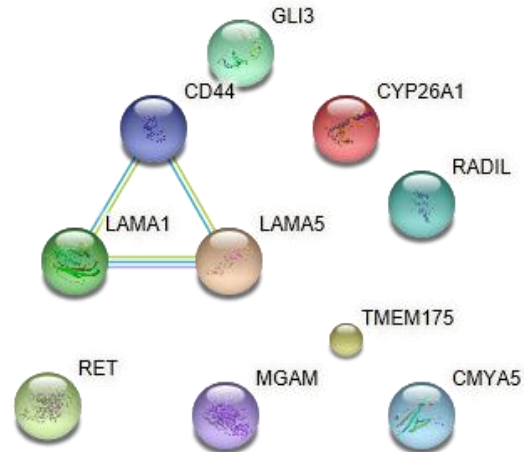
PPI enrichment p-value: 0.0211

your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

- no significant enrichment detected -

Patient HK9C



Network Stats

number of nodes: 10
 number of edges: 3
 average node degree: 0.6
 clustering coefficient: 1

expected number of edges: 0
 PPI enrichment p-value: 0.014

your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

Biological Process (GO)

pathway ID	pathway description	count in gene set	false discovery rate
GO:0001657	ureteric bud development	4	0.000272
GO:0001823	mesonephros development	4	0.000272
GO:0022612	gland morphogenesis	4	0.000371
GO:0060562	epithelial tube morphogenesis	5	0.000376
GO:0048754	branching morphogenesis of an epithelial tube	4	0.000754

(more ...)

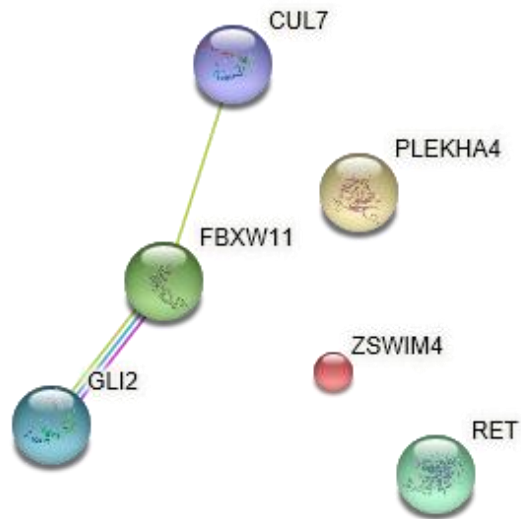
Cellular Component (GO)

pathway ID	pathway description	count in gene set	false discovery rate
GO:0043256	laminin complex	2	0.00984
GO:0005605	basal lamina	2	0.0404

KEGG Pathways

pathway ID	pathway description	count in gene set	false discovery rate
04512	ECM-receptor interaction	3	0.00164
05200	Pathways in cancer	4	0.00164

Patient VH105C



Network Stats

number of nodes: 6
number of edges: 2
average node degree: 0.667
clustering coefficient: 0.833

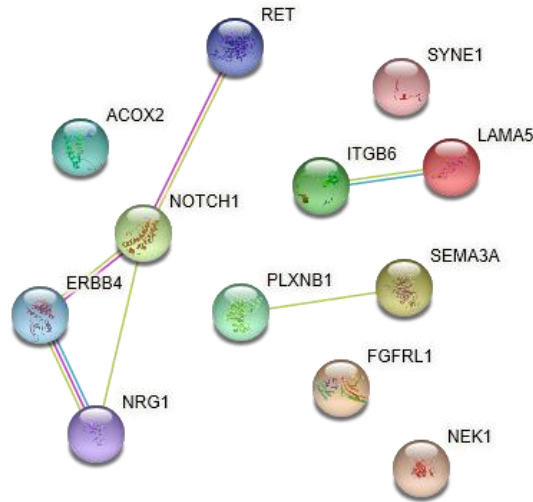
expected number of edges: 0
PPI enrichment p-value: 0.00828
your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

KEGG Pathways

pathway ID	pathway description	count in gene set	false discovery rate
04340	Hedgehog signaling pathway	2	0.025

Patient VH106C



Network Stats

number of nodes: 12
 number of edges: 6
 average node degree: 1
 clustering coefficient: 0.944

expected number of edges: 1
 PPI enrichment p-value: 0.00185
your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

Biological Process (GO)

pathway ID	pathway description	count in gene set	false discovery rate
GO:0014031	mesenchymal cell development	6	1.46e-07
GO:0048762	mesenchymal cell differentiation	6	1.46e-07
GO:0014032	neural crest cell development	5	3.13e-07
GO:0014033	neural crest cell differentiation	5	4.92e-07
GO:0060485	mesenchyme development	6	7.11e-07

(more ...)

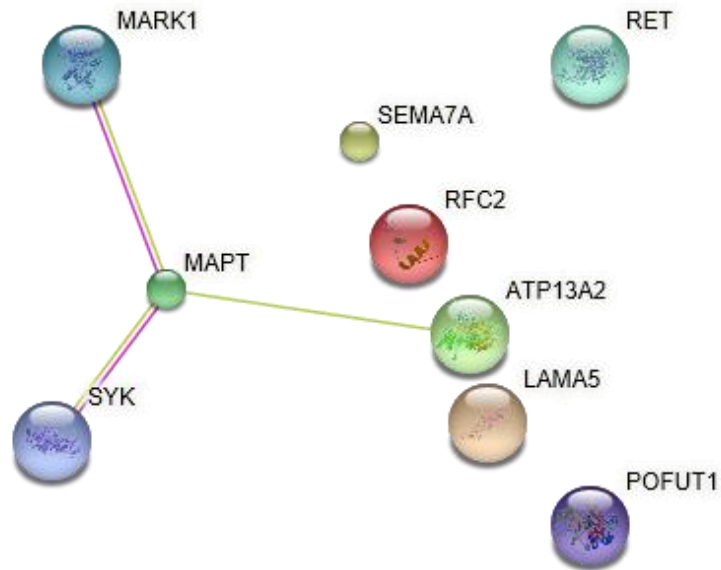
Molecular Function (GO)

pathway ID	pathway description	count in gene set	false discovery rate
GO:0004714	transmembrane receptor protein tyrosine kinase activity	3	0.0164
GO:0005102	receptor binding	6	0.0164
GO:0004713	protein tyrosine kinase activity	3	0.0456
GO:0004872	receptor activity	6	0.0456

Cellular Component (GO)

pathway ID	pathway description	count in gene set	false discovery rate
GO:0043235	receptor complex	5	0.000508

Patient VH108C



Network Stats

number of nodes: 9
number of edges: 3
average node degree: 0.667
clustering coefficient: 0.889

expected number of edges: 0
PPI enrichment p-value: 0.0114

your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

- no significant enrichment detected -

Patients	Gene	Variant	Reason for gene selection	Type	Gene function	Human	Mouse	
VH106C	SYNE1	p.S2126F/p.R5617*	Seed	CH	Spectrin family rSYNE1-RelatSYNE2 is required fo			
	ERBB4	c.1125-3178delAAACAG (intron 9)	ENS-gene	NCDS de novo	NRG1 receptor. ErbB4-Relat	Abnormal multiorga		
	NRG1	c.746-294155G>C (intron 1)	ENS-gene	NCDS de novo	Direct ligand forAssociated	Abnormal NCC mig		
	NOTCH1	p.R1330H (cu)	ENS-gene	Paternally inherited	Transmembranc	Adams-Oliv	Abnormal embryon	
	ACOX2	p.R88Q	Interacting partner; Seed	-CH in patient HK180-	Maternally inherited	Acyl-CoA oxida	ND	ND
	FGFRL1	p.V98L	Seed	-CH in patient HD9-	Paternally inherited	Fibroblast Grow	Wolf-Hirschh	Development dela
	NEK1	p.R355G	Seed	-Homo in patient HK164-	Paternally inherited	Cell cycle regul	Short-rib tho	Abnormal renal tub
	LAMA5	p.L2185F	Seed	- CH in patient HK9C-	Paternally inherited	Extracellular mc	ND	Arrest of hair devel
	ITGB6	p.R499H	Interacting partner of LAMA5		Maternally inherited	Integrin. Attach:	Amelogene	ND
	PLXNB1	p.L1686M	ENS-gene		Maternally inherited	Semaphorin rec	ND	ND
SEMA3A	c.1453-3693TC>-C (intron 12)	Seed + ENS gene		NCDS de novo	Involved in the Hypogonad	Abnormal ENS morp		