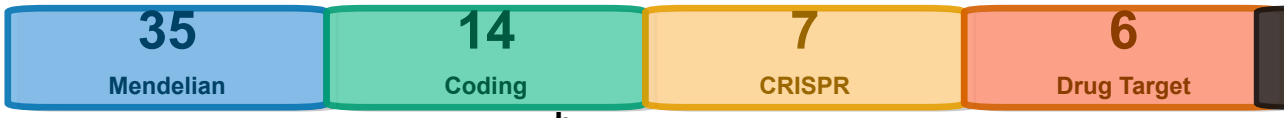
Extended Data Table 1: Post-2021 Independent Benchmark Genes

63 genes from publications post-dating L2G training cutoff (2021) · Verified absent from all training sets



a

b

Tier1_Mendelian · 35 genes

ClinGen Definitive evidence · (Mendelian disease causative)

LDLR	APOB	**APOE**
BRCA1	TP53	**CFTR**
GCK	HNF1A	KCNJ11
MC4R	**HBB**	HBA1
DMD	PKD1	UMOD
APC	SNCA	LRRK2
GBA	MYBPC3	SERPINA1
F8	FMR1	HFE
LEP	LRP5	COL1A1
GJB2	SLC26A4	MYOC
ABCA4	HEXA	HLA-DRB1
TSHR	TG	

Tier1_Coding · 14 genes

Functional coding variants · (protein-altering mutations)

LPA CAD/Lp(a) levels	**SLC30A8** T2D protection (LoF)
APOC3 Triglycerides/CAD	**NOD2** Crohn's disease
PTPN22 RA/T1D R620W	**TYK2** SLE/MS P1104A
APOL1 CKD G1/G2 variants	**SLC2A9** Uric acid/gout
PNPLA3 NAFLD I148M	**TM6SF2** NAFLD E167K
CFH AMD Y402H	**TREM2** AD R47H
SORL1 AD trafficking	**IRS1** T2D insulin resistance

c

d

Tier1_CRISPR · 7 genes

CRISPR/perturbation validated · (functional screen evidence)

PCSK9 LDL/CAD drug target	**ANGPTL3** LDL/TG evinacumab	**IRX3** Obesity FTO target
CHD8 Autism chromatin	**WNT16** Bone density Wnt	**MTNR1B** Glucose/circadian

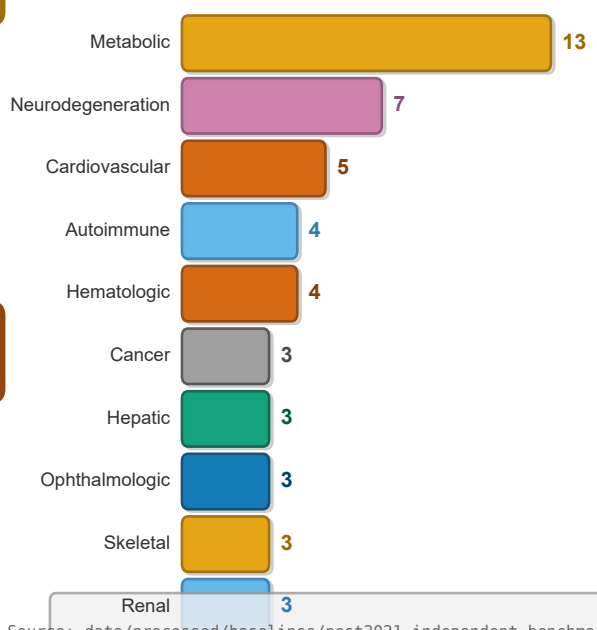
Tier1_Drug · 6 genes

FDA-approved targets · (2021-2024 approvals)

IL23R IBD anti-IL23	**PPARG** T2D thiazolidinediones	**JAK2** MPN JAK inhibitors
CETP HDL/CAD inhibitors	**CACNA1C** Psych Ca channel	

Disease Categories · 17 areas covered

Comprehensive coverage across human disease



Source: data/processed/baselines/post2021_independent_benchmark_FINAL.tsv