Open-Source Project

Intelligence Report

HCLS AI Factory — Genomics to Drug Discovery

Pipeline Run: HCLS-VCP-2026-0087

Field	Value
Patient ID	GEN-2026-0087
Run ID	HCLS-VCP-2026-0087
Pipeline Version	HLS-Pipeline v1.0.0
Pipeline Mode	Full (Genomics → RAG/Chat → Drug Discovery)
Hardware	NVIDIA DGX Spark (GB10 GPU, 128 GB unified)
Total Duration	4 hours 12 minutes
Report Date	February 2026
Status	COMPLETE — 100 novel drug candidates ranked

PIPELINE COMPLETE | **PRIMARY TARGET: VCP (Pathogenic)** | **100 CANDIDATES RANKED**

1. Genomics Summary (Stage 1)

Input Data

Parameter	Value
Sample	HG002 (NA24385, GIAB Ashkenazi male)
Sequencing	Illumina, 30× WGS, 2×250 bp paired-end
FASTQ Size	198.7 GB (R1: 99.4 GB, R2: 99.3 GB)
Reference	GRCh38 (3.1 GB)

Parabricks Execution

Step	Tool	Duration	GPU Util	Peak Memory
Alignment	BWA-MEM2 (fq2bam)	34 min	82%	38 GB
Variant Calling	Google DeepVariant	22 min	91%	54 GB
Total Stage 1		56 min		

VCF Output

Metric	Count
Total Variants Called	11,724,891
PASS Quality (QUAL>30)	3,487,216
SNPs	4,198,433
Indels	1,012,548
Multi-allelic Sites	148,762
Coding Region Variants	35,616
Ts/Tv Ratio	2.07

Quality Assessment: **PASS** — Ts/Tv ratio within expected range (2.0-2.1), variant counts consistent with 30× WGS of Ashkenazi ancestry sample.

2. Annotation & Target Identification (Stage 2)

Annotation Funnel

Stage	Variants	Filter Applied
Raw VCF	11,724,891	—
Quality Filter	3,487,216	QUAL > 30
ClinVar Annotated	35,616	Clinical significance match
AlphaMissense Scored	6,831	AI pathogenicity prediction
HIGH Impact + Pathogenic	2,412	Actionable subset
Druggable Gene Targets	847	Knowledge base match

Top 5 Target Hypotheses (Claude RAG Analysis)

Rank	Gene	Variant	ClinVar	AM Score	Area	Druggability
1	VCP	rs188935092	Pathogenic	0.87	Neurology	0.92
2	EGFR	rs121913229	Pathogenic	0.79	Oncology	0.95
3	BRCA1	rs80357914	Pathogenic	0.72	Oncology	0.78
4	PCSK9	rs11591147	Pathogenic	0.68	Cardiovascular	0.88
5	CFTR	rs75527207	Pathogenic	0.81	Respiratory	0.71

Primary Target: VCP

Parameter	Value
Gene	VCP (Valosin-Containing Protein / p97)
UniProt	P55072
Function	AAA+ ATPase, ubiquitin-proteasome pathway
Diseases	Frontotemporal Dementia (FTD), ALS, IBMPFD
Variant	rs188935092 — missense, HIGH impact
ClinVar	Pathogenic (reviewed by expert panel)
AlphaMissense	0.87 (pathogenic, >0.564 threshold)
Druggability	0.92 (D2 ATPase domain, ~450 Å^3)
Known Inhibitors	CB-5083 (Phase I), NMS-873
Confidence	HIGH — multiple independent evidence sources

Evidence Chain

1. Genomic: rs188935092 at chr9:35065263 (G>A), heterozygous, QUAL=892
2. Clinical: ClinVar Pathogenic for FTD/ALS/IBMPFD (expert panel)
3. AI Prediction: AlphaMissense 0.87 (>0.564 pathogenic threshold)
4. Functional: VEP missense_variant, HIGH impact, D2 ATPase domain
5. Druggability: Known target — CB-5083 reached Phase I clinical trial
6. Structural: 4 PDB structures including inhibitor-bound 5FTK

3. Drug Discovery Results (Stage 3)

Structure Evidence

PDB ID	Resolution	Method	Description	Score
5FTK	2.3 Å	X-ray	VCP D2 + CB-5083 inhibitor	13.2 (selected)
7K56	2.5 Å	Cryo-EM	VCP complex	10.8
8OOI	2.9 Å	Cryo-EM	WT VCP hexamer	8.9
9DIL	3.2 Å	Cryo-EM	Mutant VCP	7.4

Selected: 5FTK — inhibitor-bound (CB-5083), X-ray at 2.3 Å. Binding site: D2 ATPase domain, key residues ALA464, GLY479, ASP320, GLY215.

Molecule Generation (MolMIM)

Parameter	Value
Seed Compound	CB-5083 (ATP-competitive VCP inhibitor)
NIM Endpoint	MolMIM (port 8001)
Molecules Generated	100
Chemically Valid	98 (2 rejected by RDKit)
Generation Time	2 min 14 sec

Drug-Likeness Profile

Metric	Pass	Fail	Pass Rate
Lipinski Rule of Five	87	11	88.8%
QED > 0.67 (drug-like)	72	26	73.5%
QED > 0.49 (moderate+)	91	7	92.9%
TPSA < 140 Å²	94	4	95.9%

Molecular Docking (DiffDock)

Parameter	Value
NIM Endpoint	DiffDock (port 8002)
Protein Target	5FTK (VCP D2 domain)
Candidates Docked	98
Docking Time	8 min 42 sec
Mean Dock Score	-7.4 kcal/mol
Best Dock Score	-11.4 kcal/mol
Excellent (< -8.0)	34 candidates
Good+ (< -6.0)	78 candidates

Top 10 Ranked Candidates

Composite scoring: 30% Generation + 40% Docking + 30% QED

Rank	Composite	Gen	Dock	QED	MW	LogP	Lipinski
1	0.89	0.92	-11.4	0.81	423.5	3.2	PASS
2	0.86	0.88	-10.8	0.79	441.2	3.7	PASS
3	0.84	0.85	-10.2	0.82	398.7	2.9	PASS
4	0.82	0.91	-9.8	0.74	467.1	4.1	PASS
5	0.81	0.83	-9.5	0.78	412.3	3.4	PASS
6	0.79	0.87	-9.1	0.71	455.8	3.8	PASS
7	0.78	0.80	-8.9	0.76	389.2	2.7	PASS
8	0.76	0.84	-8.7	0.69	478.4	4.3	PASS
9	0.75	0.79	-8.5	0.73	401.6	3.1	PASS
10	0.74	0.82	-8.2	0.68	448.9	3.9	PASS

CB-5083 Seed Comparison

Metric	CB-5083 (Seed)	Top Candidate	Improvement
Dock Score	-8.1 kcal/mol	-11.4 kcal/mol	+41% binding
QED	0.62	0.81	+31% drug-likeness
MW	487.2 Da	423.5 Da	-13% (better absorption)
Composite	0.64	0.89	+39% overall

4. Pipeline Performance

Stage Timing

Stage	Duration	GPU Util	Peak Memory
1 — Genomics (fq2bam)	34 min	82%	38 GB
1 — Genomics (DeepVariant)	22 min	91%	54 GB
2 — Annotation	18 min	15% (CPU)	12 GB
2 — Milvus Indexing	24 min	35%	22 GB
2 — RAG/Chat	45 min	5%	8 GB
3 — Structure Retrieval	2 min	0% (I/O)	2 GB
3 — MolMIM Generation	2 min 14 sec	78%	18 GB
3 — DiffDock Docking	8 min 42 sec	85%	24 GB
3 — Scoring + Reporting	1 min 30 sec	0% (CPU)	4 GB
Total	~4 hr 12 min		

All Services Healthy

Service	Port	Status
Landing Page	8080	HEALTHY
Genomics Portal	5000	HEALTHY
Milvus	19530	HEALTHY
RAG API	5001	HEALTHY
Streamlit Chat	8501	HEALTHY
MolMIM NIM	8001	HEALTHY
DiffDock NIM	8002	HEALTHY
Discovery UI	8505	HEALTHY
Grafana	3000	HEALTHY
Prometheus	9099	HEALTHY

5. Clinical Interpretation

Summary

Patient GEN-2026-0087 carries a heterozygous pathogenic missense variant (rs188935092) in the VCP gene. This variant is associated with Frontotemporal Dementia (FTD), ALS, and Inclusion Body Myopathy with Paget Disease and Frontotemporal Dementia (IBMPFD). The variant is classified as Pathogenic by ClinVar expert panel review and scores 0.87 on the AlphaMissense pathogenicity scale.

Drug Discovery Outcome

The AI-driven drug discovery pipeline identified 100 novel VCP inhibitor candidates with the top candidate showing a 39% improvement in composite score over the CB-5083 seed compound. All top 10 candidates pass Lipinski's Rule of Five and show favorable QED scores (>0.67), suggesting oral drug-likeness.

Recommended Actions

1. Genetic counseling for FTD/ALS risk assessment
2. Experimental validation of top 5 candidates in VCP ATPase assays
3. ADMET profiling for lead optimization
4. Cross-modal follow-up with Imaging Intelligence Agent for neurological assessment

6. Provenance

Item	Value
Pipeline	HLS-Pipeline v1.0.0 (Nextflow DSL2)
Parabricks	nvcr.io/nvidia/clara/clara-parabricks:4.6.0-1
DeepVariant	Google DeepVariant (via Parabricks, >99%)
Reference	GRCh38 (3.1 GB)
ClinVar	February 2026 release (4.1M variants)
AlphaMissense	v1.0 (71,697,560 predictions)
VEP	Ensembl VEP (GRCh38)
Milvus	v2.4 (IVF_FLAT, nlist=1024, COSINE)
Embedding	BGE-small-en-v1.5 (384-dim)
LLM	claude-sonnet-4-20250514 (temp=0.3)
MolMIM	nvcr.io/nvidia/clara/bionemo-molmim:1.0
DiffDock	nvcr.io/nvidia/clara/diffdock:1.0
Hardware	NVIDIA DGX Spark (GB10, 128 GB)
Scoring	30% gen + 40% dock + 30% QED

This is a demonstration intelligence report. All patient data is synthetic.