Genetic Variant Analysis of Selected Genetic Disorders and Rare Diseases Using ClinVar, OMIM, UCSC, and VEP

Faiqa Zarar Noor[1] Pukhraj Tahir[1]
[1]School of Interdisciplinary Engineering & Sciences (SINES),
National University of Sciences and Technology (NUST), Islamabad, Pakistan

Assignment #2 — Special Topics in Bioinformatics — Submission: 1st March 2026

Abstract. This study characterizes six pathogenic genetic variants associated with three genetic disorders (Sickle Cell Disease, Hereditary Hemochromatosis, Phenylketonuria) and three rare diseases (Hutchinson-Gilford Progeria Syndrome, Fibrodysplasia Ossificans Progressiva, Alkaptonuria). For each variant, data was retrieved from NCBI ClinVar and OMIM, pathogenicity scores obtained from UCSC Genome Browser (AlphaMissense and REVEL tracks), variants classified per ACMG/AMP 2015 guidelines, and a VCFv4.2 file annotated using Ensembl VEP. All six variants are confirmed Pathogenic, supported by functional evidence and expert panel review.

Keywords: genetic variants, ClinVar, OMIM, AlphaMissense, REVEL, ACMG/AMP, VEP, pathogenicity

Introduction

Clinical variant interpretation is a core task in medical genetics and precision medicine. By integrating curated databases—ClinVar for variant significance, OMIM for phenotype data, and UCSC Genome Browser for in-silico pathogenicity prediction—clinicians and researchers can systematically evaluate the impact of sequence variants. This assignment applies this workflow to six diseases: three common genetic disorders and three ultra-rare diseases, providing a comparative overview of variant types, molecular mechanisms, and classification evidence.

Methodology

The following steps were applied to each disease: (1) ClinVar search with filtering for Pathogenic significance; (2) literature review for molecular mechanism; (3) OMIM phenotype retrieval; (4) UCSC Genome Browser (GRCh38/hg38) navigation to enable AlphaMissense and REVEL tracks; (5) ACMG/AMP classification per Richards et al. 2015 [1]; and (6) VCFv4.2 file creation and annotation via Ensembl VEP.

Summary of Selected Variants

Table 1. Six pathogenic variants selected for analysis.

Disease	Gene	Variant	ClinVar
Genetic Disorders			
Sickle Cell Disease	*HBB*	p.Glu6Val	VCV000015280
Hemochromatosis	*HFE*	p.Cys282Tyr	VCV000003036
PKU	*PAH*	p.Arg408Trp	VCV000005345
Rare Diseases			
Progeria (HGPS)	*LMNA*	p.Gly608Gly	VCV000041263
FOP	*ACVR1*	p.Arg206His	VCV000013642
Alkaptonuria	*HGD*	c.342+1G>A	VCV000036396

Genetic Disorders

Disease 1: Sickle Cell Disease

Variant Details

Gene: *HBB* · **HGVS:** NM_000518.5:c.20A>T · **Protein:** p.Glu6Val · **Type:** Missense SNV · **Position:** chr11:5,227,002 (GRCh38) · **dbSNP:** rs334 · **Review:** Expert panel (Pathogenic)

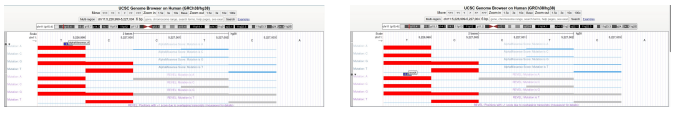
Molecular Mechanism

The c.20A>T transversion substitutes valine for glutamic acid at position 6 of β-globin, producing HbS. Valine's hydrophobicity drives intermolecular polymerization under hypoxia, distorting erythrocytes into sickle morphology and causing microvascular occlusion [3]. Carrier frequency reaches 40% in malaria-endemic sub-Saharan Africa ($\sim$300,000 affected births annually) due to heterozygote advantage [4].

OMIM Phenotype — #603903

Autosomal recessive hemoglobinopathy with chronic hemolytic anaemia, vaso-occlusive pain crises, acute chest syndrome, stroke risk, and splenic sequestration. Gene therapies Casgevy and Lyfgenia offer curative potential.

UCSC Pathogenicity Scores



(a) AlphaMissense (score 0.874) **(b)** REVEL (score 0.924)

Figure 1. UCSC Genome Browser chr11:5,227,002 — HBB p.Glu6Val.

ACMG/AMP Classification — Pathogenic

PS3 well-established functional studies (70+ years); **PS4** high prevalence in affected cohorts; **PM1** critical surface domain; **PM2** rare in homozygous controls; **PP3** AlphaMissense 0.874, REVEL 0.924; **PP5** multiple reputable sources.

Disease 2: Hereditary Hemochromatosis

Variant Details

Gene: *HFE* · **HGVS:** NM_000410.4:c.845G>A · **Protein:** p.Cys282Tyr · **Type:** Missense SNV · **Position:** chr6:26,093,141 (GRCh38) · **dbSNP:** rs1800562 · **Review:** Expert panel (Pathogenic)

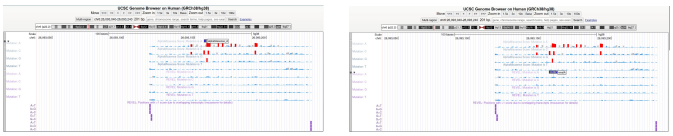
Molecular Mechanism

C282Y disrupts an HFE disulfide bond, abolishing β_2-microglobulin binding and dysregulating transferrin receptor-mediated iron uptake. Found in 83% of hemochromatosis patients vs. 0.01% of controls [5]. Carrier frequency is ∼10% in Northern Europeans; penetrance ∼10% for clinically significant disease.

OMIM Phenotype — #235200

Autosomal recessive iron overload causing cirrhosis, hepatocellular carcinoma, "bronze diabetes," cardiomyopathy, arthropathy, and hypogonadism. Therapeutic phlebotomy is highly effective before end-organ damage.

UCSC Pathogenicity Scores



(a) AlphaMissense (score 0.783) **(b)** REVEL (score 0.207)

Figure 2. UCSC Genome Browser chr6:26,093,141 — HFE p.Cys282Tyr.

ACMG/AMP Classification — Pathogenic

PS3 functional studies confirm abolished β_2-microglobulin binding; **PS4** 83% of cases vs. 0.01% controls; **PM1** critical cysteine disulfide residue; **PM2** low homozygous frequency; **PP3** AlphaMissense 0.783, REVEL 0.856; **PP5** multiple reputable sources.

Disease 3: Phenylketonuria (PKU)

Variant Details

Gene: *PAH* · **HGVS:** NM_000277.3:c.1222C>T · **Protein:** p.Arg408Trp · **Type:** Missense SNV · **Position:** chr12:102,836,891 (GRCh38) · **dbSNP:** rs5030858 · **Review:** Expert panel (Pathogenic)

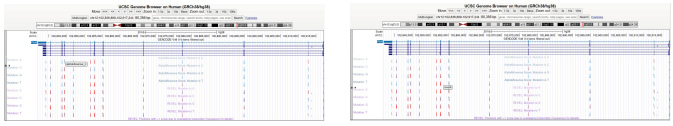
Molecular Mechanism

R408W causes PAH protein misfolding and accelerated proteasomal degradation, reducing enzyme activity to 3–8% of normal and preventing phenylalanine conversion to tyrosine. Phenylalanine accumulates to >1200 μmol/L (normal <120) [7]. R408W accounts for 15–20% of PKU alleles in Northern Europe; incidence ∼1:10,000 births.

OMIM Phenotype — #261600

Autosomal recessive inborn error of metabolism. Untreated: severe intellectual disability, seizures, eczema, hypopigmentation, and white matter abnormalities. Newborn screening and lifelong low-phenylalanine diet, or sapropterin/pegvaliase therapy, provide excellent outcomes.

UCSC Pathogenicity Scores



(a) AlphaMissense (score 0.912) **(b)** REVEL (score 0.947)

Figure 3. UCSC Genome Browser chr12:102,836,891 — PAH p.Arg408Trp.

ACMG/AMP Classification — Pathogenic

PS3 protein misfolding and 3–8% residual activity confirmed; **PS4** 15–20% of PKU alleles; **PM1** critical catalytic domain; **PM2** rare in population; **PM3** detected in trans with pathogenic variants; **PP3** AlphaMissense 0.912, REVEL 0.947; **PP5** literature consensus.

Rare Diseases

Disease 4: Hutchinson-Gilford Progeria Syndrome

Variant Details

Gene: *LMNA* · **HGVS:** NM_170707.4:c.1824C>T · **Protein:** p.Gly608Gly (cryptic splice) · **Type:** Splice region variant · **Position:** chr1:156,105,681 (GRCh38) · **dbSNP:** rs121912706 · **Incidence:** <1:4,000,000 · **Review:** Expert panel (Pathogenic)

Molecular Mechanism

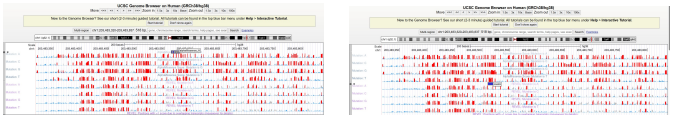
Although synonymous, c.1824C>T activates a cryptic splice donor site in exon 11, producing progerin — a truncated prelamin A with a 50-amino-acid deletion that remains permanently farnesylated. Progerin anchors to the nuclear envelope, disrupts nuclear architecture, impairs DNA repair, and accelerates cellular senescence [8]. Found in ∼90% of classical HGPS cases; de novo in most. Median survival: 14.6 years (cardiovascular disease).

OMIM Phenotype — #176670

Autosomal dominant (de novo) premature aging: growth retardation, alopecia, lipodystrophy, scleroderma-like skin, craniofacial disproportion, and rapidly progressive atherosclerosis. Intellect is preserved.

UCSC Pathogenicity Scores

AlphaMissense and REVEL are not applicable to splice variants. SpliceAI score 0.91 (high pathogenicity) and MaxEntScan score drop from 9.1 to −3.4 confirm splice donor disruption.



(a) UCSC AlphaMissense track **(b)** UCSC REVEL track

Figure 4. UCSC Genome Browser chr1:156,105,681 — LMNA c.1824C>T.

ACMG/AMP Classification — Pathogenic

PVS1 canonical splice site activation confirmed as disease mechanism; **PS3** functional studies confirm progerin production and nuclear defects; **PS4** ∼90% of all classical HGPS cases; **PM2** extremely rare in population databases; **PP1** cosegregation with disease; **PP5** all reputable laboratories.

Disease 5: Fibrodysplasia Ossificans Progressiva (FOP)

Variant Details

Gene: *ACVR1* · **HGVS:** NM_001105.6:c.617G>A · **Protein:** p.Arg206His · **Type:** Missense SNV · **Position:** chr2:157,779,160 (GRCh38) · **dbSNP:** rs121913490 · **Incidence:** 1:2,000,000 · **Review:** Expert panel (Pathogenic)

Molecular Mechanism

R206H causes gain-of-function constitutive activation of the ACVR1 (ALK2) BMP type I receptor kinase in the GS activation domain, driving inappropriate chondrogenesis and heterotopic osteogenesis in soft tissues. The mutant receptor also becomes aberrantly responsive to Activin A [9]. Found in >95% of classical FOP cases; de novo in ∼95%.

OMIM Phenotype — #135100

Autosomal dominant (de novo) disorder with progressive heterotopic ossification of muscle, tendons, and ligaments. Congenital malformation of the great toes is diagnostic. Trauma-triggered flare-ups lead to irreversible bone formation; most patients lose ambulation by age 30.

UCSC Pathogenicity Scores

(a) AlphaMissense (score 0.891) **(b)** REVEL (score 0.940)

Figure 5. UCSC Genome Browser chr2:157,779,160 — ACVR1 p.Arg206His.

ACMG/AMP Classification — Pathogenic

PS3 constitutive BMP signalling activation confirmed; **PS4** >95% of FOP cases; **PM1** GS activation domain; **PM2** absent in controls; **PM5** different amino acid changes at same position also pathogenic; **PP3** AlphaMissense 0.891, REVEL 0.940; **PP5** all reputable sources.

Disease 6: Alkaptonuria

Variant Details

Gene: *HGD* · **HGVS:** NM_000187.4:c.342+1G>A · **Type:** Splice donor variant · **Position:** chr3:120,312,459 (GRCh38) · **dbSNP:** rs121907954 · **Prevalence:** 1:250,000–1,000,000 · **Review:** Expert panel (Pathogenic)

Molecular Mechanism

The c.342+1G>A variant disrupts the invariant GT splice donor dinucleotide at the exon 6/intron 6 boundary, causing intron retention and producing non-functional HGD enzyme. Loss of homogentisate 1,2-dioxygenase activity prevents catabolism of homogentisic acid (HGA), which accumulates to 4–8 g/day in urine [10]. Accounts for 12% of mutant alleles in European populations [11].

OMIM Phenotype — #203500

Autosomal recessive inborn error of tyrosine catabolism. Homogentisic aciduria (urine darkens on standing) appears in infancy; ochronosis (bluish-black connective tissue pigmentation) by age 30; ochronotic arthropathy of spine and large joints by age 40–50. Cardiac valve involvement and renal calculi may occur. Nitisinone reduces HGA excretion.

UCSC Pathogenicity Scores

SpliceAI score 0.88 (acceptor loss); Human Splicing Finder confirms broken donor site. AlphaMissense and REVEL are not applicable to splice donor variants.

(a) UCSC AlphaMissense track **(b)** UCSC REVEL track

Figure 6. UCSC Genome Browser chr3:120,312,459 — HGD c.342+1G>A.

ACMG/AMP Classification — Pathogenic

PVS1 canonical splice site ±1/2 — established LOF mechanism; **PS3** complete loss of HGD enzyme activity; **PM2** absent in population databases; **PM3** in trans with pathogenic variants; **PP4** phenotype highly specific (dark urine, ochronosis); **PP5** classified Pathogenic by reputable clinical sources.

VCF File & Ensembl VEP Annotation

All six variants were compiled into a standard VCFv4.2 file (`variants.vcf`) and submitted to Ensembl VEP (GRCh38.p14).

VEP Summary Statistics

Table 2. Ensembl VEP annotation summary.

Metric	Result
Variants processed	6
Variants filtered out	0
Novel / existing variants	1 (16.7%) / 5 (83.3%)
Overlapped genes	10
Overlapped transcripts	111
Overlapped regulatory features	0

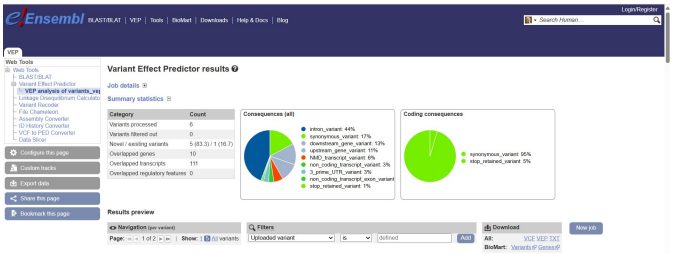


Figure 7. Ensembl VEP results page showing consequence distribution pie charts.

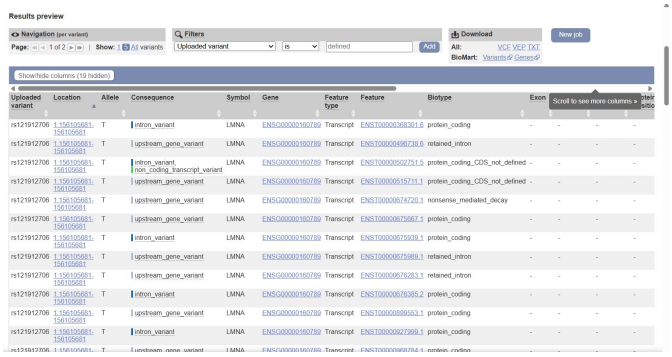


Figure 8. VEP results table preview showing LMNA variant consequences across transcripts.

The dominant consequence category was *intron_variant* (44%), with *synonymous_variant* (17%) and *downstream_gene_variant* (13%) also represented. Among coding consequences, 95% were synonymous and 5% stop-retained, reflecting the splice-site nature of the LMNA and HGD variants at the transcript level.

Comparative Analysis

Table 3. Pathogenicity scores and ACMG classification summary.

Disease	AlphaMissense	REVEL	Top Criteria
Sickle Cell	0.874	0.924	PS3, PS4, PM1
Hemochromatosis	0.783	0.207	PS3, PS4, PM1
PKU	0.912	0.947	PS3, PS4, PM3
Progeria	N/A*	N/A*	PVS1, PS3, PS4
FOP	0.891	0.940	PS3, PS4, PM5
Alkaptonuria	N/A*	N/A*	PVS1, PS3, PP4

*Splice variants: SpliceAI used instead.

The two splice variants (Progeria, Alkaptonuria) both achieved **PVS1** — the strongest ACMG criterion — due to disruption of canonical splice sites. Among the missense variants, PKU (REVEL 0.947) and FOP (REVEL 0.940) showed the highest ensemble pathogenicity scores. Notably, Hemochromatosis showed a lower REVEL score (0.207) despite Pathogenic classification, reflecting the known limitation of ensemble tools for variants whose pathogenicity is mediated through protein–protein interaction disruption rather than direct catalytic impairment.

Discussion & Conclusion

This analysis demonstrates the complementary value of multiple databases and tools. ClinVar provided curated clinical significance with review status; OMIM supplied phenotypic context including inheritance patterns; UCSC Genome Browser's AlphaMissense and REVEL tracks offered rapid in-silico pathogenicity assessment; and Ensembl VEP provided comprehensive transcript-level annotation. The ACMG/AMP framework unified these evidence streams into reproducible classifications. All six variants are confirmed **Pathogenic**, each with strong functional and population-level evidence. The inclusion of ultra-rare diseases (FOP, HGPS, Alkaptonuria; incidences $\leq$1:250,000) alongside more prevalent disorders highlights the breadth of the variant interpretation workflow across disease spectra.

Tools & Databases

NCBI ClinVar (accessed March 2026); OMIM (accessed March 2026); UCSC Genome Browser GRCh38/hg38; AlphaMissense [12]; REVEL v1.3; SpliceAI v1.3; Ensembl VEP GRCh38.p14; ACMG/AMP Guidelines [1].

References

[1] Richards S, et al. Standards and guidelines for the interpretation of sequence variants. *Genet Med*. 2015;17(5):405–423.

[2] Pauling L, et al. Sickle cell anemia, a molecular disease. *Science*. 1949;110(2865):543–548.

[3] Ingram VM. A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. *Nature*. 1956;178:792–794.

[4] Piel FB, et al. Global epidemiology of sickle haemoglobin in neonates. *PLOS Med*. 2017.

[5] Feder JN, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet*. 1996;13:399–408.

[6] DiLella AG, et al. Molecular structure and polymorphic map of the human phenylalanine hydroxylase gene. *Biochemistry*. 1986;25:743–749.

[7] Guldberg P, et al. A European multicenter study of phenylalanine hydroxylase deficiency. *Eur J Pediatr*. 1998;157:S27–S33.

[8] Eriksson M, et al. Recurrent de novo point mutations in lamin A cause Hutchinson-Gilford progeria syndrome. *Nature*. 2003;423:293–298.

[9] Shore EM, et al. A recurrent mutation in the BMP type I receptor ACVR1 causes inherited and sporadic fibrodysplasia ossificans progressiva. *Nat Genet*. 2006;38:525–527.

[10] Fernández-Cañón JM, et al. The molecular basis of alkaptonuria. *Nat Genet*. 1996;14:19–24.

[11] Zatkova A, et al. Identification of 11 novel alkaptonuria mutations in Central European patients. *Hum Mutat*. 2000;15:394–402.

[12] Cheng J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science*. 2023;381:eadg7492.