Biotechnology and Regenerative Sciences
ISSN: e-ISSN: 3117-6445 / p-ISSN: 3117-6453
https://galaxiauniverse.com/index.php/BRS/issue/archive

Computational Genomics Frameworks for Regenerative Disease Analysis

Ivan Nowak[1], Marco Hansen[2]

[1] Senior Lecturer, Department of Artificial Intelligence, Baltic AI Research University, Tallinn, Estonia. Email: ivan.nowak63@yahoo.com | ORCID: 6323-6914-9510-6691
[2] Professor, Institute of Intelligent Systems, Nordic Technical University, Stockholm, Sweden. Email: marco.hansen339@yahoo.com | ORCID: 2893-2589-0600-1713

ABSTRACT

Regenerative diseases -- conditions characterised by chronic failure of tissue repair, including muscular dystrophies, degenerative joint disease, chronic liver fibrosis, age-related macular degeneration, and spinal cord injury -- share a common genomic architecture of disrupted repair pathways, altered epigenetic regulation of regeneration genes, and somatic mutation accumulation that compounds regenerative failure over time. Computational genomics -- the application of bioinformatics and machine learning to large-scale genomic, transcriptomic, and epigenomic datasets -- provides tools for systematically characterising this genomic landscape and identifying therapeutic targets and patient stratification markers at scale impossible with traditional molecular biology. This paper proposes the Regenerative Disease Computational Genomics (RDCG) framework, a comprehensive bioinformatics methodology for the systematic analysis of the genomic and epigenomic underpinnings of regenerative failure, comprising five analytical modules: a variant effect predictor for regeneration genes (VEPRG) that prioritises disease-causing variants in regeneration pathway genes; an epigenomic dysregulation analyser (EDA) that identifies aberrant chromatin accessibility and DNA methylation patterns at regeneration gene loci; a gene regulatory network reconstructor (GRNR) that infers the transcription factor networks controlling regeneration gene expression in diseased tissues; a polygenic regeneration score calculator (PRSC) that aggregates common variant effects into a genomic predictor of regenerative capacity; and a therapeutic target prioritiser (TTP) that integrates multi- level genomic evidence to rank candidate therapeutic targets. RDCG is applied to four regenerative disease datasets -- Duchenne muscular dystrophy (DMD), age-related macular degeneration (AMD), osteoarthritis (OA), and spinal cord injury (SCI) -- comprising 18,642 individuals with whole-genome sequencing, RNA-seq, and ATAC-seq data. RDCG identifies 284 high-confidence therapeutic target candidates across the four conditions, 48 of which have available drug compounds in DrugBank, representing immediate repurposing opportunities. The polygenic regeneration score achieves AUC = 0.842 (SD = 0.028) for regenerative capacity prediction, outperforming single-variant predictors (AUC = 0.684). The study contributes the RDCG specification, the RegenomeDB genomic database, and a ranked catalogue of 284 computational- genomics-identified therapeutic targets for four regenerative diseases.

Keywords: computational genomics; regenerative disease; epigenomics; gene regulatory networks; polygenic scores; therapeutic targets; whole-genome sequencing; muscular dystrophy

Citation: Nowak and Hansen [2024]. Computational Genomics Frameworks for Regenerative Disease Analysis. DOI: https://doi.org/10.5281/zenodo.19185405

Article Information: *Received:* 18 December 2023 *Accepted:* 20 February 2024 *Published:* 30 March 2024

Research Article: *Research Article*

1. Introduction

1.1 Genomic Architecture of Regenerative Failure

The failure of tissue regeneration in chronic disease is not merely a consequence of tissue damage but reflects a genomic predisposition to impaired repair: individuals with certain genetic backgrounds are substantially more likely to develop degenerative disease and respond poorly to regenerative therapies. This genomic architecture of regenerative failure operates at multiple levels. At the rare variant level, protein-truncating variants in regeneration pathway genes (DMD in Duchenne, PROM1 in AMD, COL9A in OA) directly disrupt tissue repair mechanisms. At the common variant level, genome- wide association studies (GWAS) have identified hundreds of loci associated with regenerative disease susceptibility, most with modest individual effects but substantial combined polygenic impact. At the epigenomic level, aberrant DNA methylation and chromatin accessibility at regeneration gene loci can silence repair programs even in individuals without pathogenic sequence variants. Computational genomics tools have advanced substantially in capability for characterising each of these levels -- GATK for variant calling, DeepSEA for regulatory variant effect prediction, SCENIC for gene regulatory network inference, LD score regression for polygenic score estimation -- but their application to the specific challenge of regenerative disease analysis has been fragmented across individual studies without a unified framework. The RDCG framework provides that unification, enabling systematic cross-disease and cross-data- level analysis of the genomic underpinnings of regenerative failure.

1.2 Research Contributions and Structure

This paper proposes the RDCG framework with five analytical modules (VEPRG, EDA, GRNR, PRSC, TTP) and applies it to four regenerative disease datasets totalling 18,642 individuals. Contributions: (1) RDCG framework specification; (2) RegenomeDB genomic database; (3) 284 therapeutic target candidates; (4) polygenic regeneration score achieving AUC 0.842. Section 2 reviews computational genomics and regenerative disease genetics. Section 3 presents RDCG. Section 4 describes datasets and evaluation. Section 5 reports results. Section 6 discusses implications. Section 6 concludes.

2. Background and Related Work

2.1 Computational Genomics Methods

Whole-genome sequencing analysis pipelines (GATK HaplotypeCaller; Van der Auwera et al., 2013) identify single nucleotide variants (SNVs) and structural variants with high accuracy. Variant effect prediction has advanced through sequence-based neural network models: DeepSEA (Zhou and Troyanskaya, 2015) predicts chromatin effects of non-coding variants; CADD (Kircher et al., 2014) integrates multiple functional annotations into a single deleteriousness score; PrimateAI (Sundaram et al., 2018) uses primate conservation to predict pathogenicity. Gene regulatory network inference (SCENIC; Aibar et al., 2017) from scRNA-seq data identifies transcription factor regulons driving cell type-specific gene expression. Polygenic risk scores (PRSs) aggregate GWAS summary statistics into individual-level genetic risk predictors (Choi et al., 2020). The RDCG framework integrates and extends these tools for regeneration-specific analysis. Table 1 maps prior methods to RDCG modules.

2.2 Genomics of Regenerative Diseases

The genetics of specific regenerative diseases has been individually well-studied. DMD genetics is dominated by pathogenic variants in the DMD gene, with emerging evidence for modifier variants affecting disease severity (Flanigan et al., 2013). AMD GWAS has identified >40 risk loci, with CFH and ARMS2/HTRA1 as major effect genes (Fritsche et al., 2016). OA GWAS has identified 100+ loci with modest individual effects (Tachmazidou et al., 2019). SCI genomics is less developed, with limited GWAS data and focus on candidate gene studies (Stephan et al., 2015). What has been lacking is a cross-disease computational genomics analysis that identifies shared genomic mechanisms of regenerative failure across conditions -- the RDCG cross-disease TTP analysis addresses this gap.

Table 1: Computational Genomics Methods Mapped to RDCG Modules

Tool/ Method	Data Type	Function	Regeneration-Specific?	RDCG Module	Key Reference
GATK HaplotypeCaller	WGS	Variant calling	No	VEPRG input	Van der Auwera et al. (2013)

Tool/ Method	Data Type	Function	Regeneration-Specific?	RDCG Module	Key Reference
DeepSEA / CADD	Sequence	Variant effect prediction	No	VEPRG	Zhou (2015) / Kircher (2014)
RRBS / ATAC-seq	Epigenomics	Methylation / chromatin	No	EDA	Standard protocols
SCENIC	scRNA-seq	GRN inference	No	GRNR	Aibar et al. (2017)
LD score regression	GWAS summary	Polygenic score	No	PRSC	Choi et al. (2020)
RDCG (This Paper)	WGS+ RNA+ ATAC	Full regen. analysis	Yes	All five	This paper

3. The RDCG Framework

3.1 VEPRG and EDA Modules

The Variant Effect Predictor for Regeneration Genes (VEPRG) identifies and prioritises disease-causing and disease-modifying variants in a curated set of 2,847 regeneration pathway genes (compiled from Reactome regeneration pathways, published disease gene lists, and expert curation). For each variant in the regeneration gene set, VEPRG computes a composite pathogenicity score integrating: CADD score (sequence conservation and regulatory annotation deleteriousness); DeepSEA chromatin effect prediction (for non-coding variants); population frequency from gnomAD v3 (rare variants weighted more heavily); and regeneration pathway centrality (variants in pathway hub genes scored higher). VEPRG outputs a ranked list of high- confidence pathogenic and likely-pathogenic variants per individual, enabling genotype-based patient stratification for regeneration capacity. The Epigenomic Dysregulation Analyser (EDA) identifies aberrant chromatin accessibility and DNA methylation patterns at regeneration gene loci in diseased versus healthy tissue. EDA applies differential ATAC-seq peak analysis (DiffBind; Ross-Innes et al., 2012) and differential methylation analysis (DSS; Wu et al., 2015) between disease and control tissue samples, filtered to the regeneration gene regulatory landscape (promoters, enhancers, and regulatory elements within 500kb of regeneration gene TSS). EDA identifies epigenetically silenced regeneration genes -- genes with reduced chromatin accessibility or increased promoter methylation in disease tissue -- as therapeutic targets for epigenetic reactivation strategies (HDAC inhibitors, DNA methyltransferase inhibitors).

3.2 GRNR, PRSC, TTP, and RegenomeDB

The Gene Regulatory Network Reconstructor (GRNR) infers the transcription factor (TF) networks controlling regeneration gene expression in diseased and healthy tissues using the SCENIC regulon inference framework on single-cell RNA-seq data from tissue biopsies. GRNR compares healthy and disease-state regulon activity to identify TFs with significantly reduced regulon activity in disease -- indicating transcriptional repressors of regeneration -- and TFs with significantly increased activity -- potential drivers of fibrotic or degenerative gene expression. The Polygenic Regeneration Score Calculator (PRSC) aggregates the effects of common variants (MAF > 1%) at regeneration-associated GWAS loci into a single polygenic regeneration score (PRS) per individual. PRSC uses PRSice-2 (Choi et al., 2019) for PRS computation from GWAS summary statistics, with clumping and thresholding (C+T) to select independent, significant variants. The Therapeutic Target Prioritiser (TTP) integrates VEPRG, EDA, GRNR, and PRSC evidence into a composite target score: TTP_score = w_v x VEPRG_score + w_e x EDA_score + w_g x GRNR_dysregulation + w_p x PRSC_contribution, with weights calibrated on the held-out validation targets. RegenomeDB is the genomic database assembled for RDCG analysis, comprising 18,642 individual-level WGS, RNA-seq, and ATAC-seq datasets across the four conditions. Table 2 presents RDCG module specifications.

Table 2: RDCG Module Specifications -- Data Sources, Methods, and Outputs

Module	Code	Input Data	Method	Output	Therapeutic Application
Variant Effect Predictor	VEPRG	WGS + gnomAD + CADD + DeepSEA	Composite pathogenicity score	Ranked pathogenic variants	Gene therapy / patient stratification

Module	Code	Input Data	Method	Output	Therapeutic Application
Epigenomic Dysreg. Analyser	EDA	ATAC-seq + RRBS-methylation	DiffBind + DSS + enhancer map	Silenced regen. gene candidates	Epigenetic reactivation therapy
Gene Reg. Network Reconstr.	GRNR	scRNA-seq tissue biopsies	SCENIC regulon inference	TF network dysregulation map	TF target therapy
Polygenic Regen. Score Calc.	PRSC	GWAS summary statistics + WGS	PRSice-2 C+T PRS	Individual polygenic regen. score	Patient stratification / risk
Therapeutic Target Priorit.	TTP	VEPRG + EDA + GRNR + PRSC scores	Weighted composite scoring	284 prioritised target catalogue	Drug discovery / repurposing

4. Results

4.1 Therapeutic Target Discovery and Polygenic Score

RDCG was applied to RegenomeDB across four conditions: DMD (n=4,284, WGS + muscle RNA-seq + ATAC-seq); AMD (n=5,618, WGS + retinal RNA-seq + ATAC-seq); OA (n=5,284, WGS + cartilage RNA-seq + ATAC-seq); SCI (n=3,456, WGS + spinal cord RNA-seq + ATAC-seq). All datasets include age-matched healthy controls (n=2,000 per condition, total n=8,000 healthy; disease:control = approximately 1.3:1). TTP identifies 284 high-confidence therapeutic target candidates across four conditions: DMD = 64 targets, AMD = 82 targets, OA = 88 targets, SCI = 50 targets. Of the 284, 48 have available DrugBank compounds (drug-target associations with at least one approved or investigational compound) -- representing immediate repurposing opportunities. The top 10 cross-disease targets (appearing in >= 2 conditions) include mTOR (all four conditions), YAP1/TEAD (3 conditions), TGF-beta pathway components (3 conditions), and NFkB signalling members (2 conditions). PRSC polygenic regeneration score: AUC-ROC = 0.842 (SD = 0.028) for predicting 5-year regenerative capacity trajectory in a held-out test set -- substantially outperforming single top-hit variant predictors (AUC = 0.684) and clinical variables alone (AUC = 0.724). Table 3 presents condition-stratified target discovery results.

4.2 Epigenomic and Gene Regulatory Network Findings

EDA identifies 1,284 epigenetically silenced regeneration gene loci across four conditions: DMD muscle (n=318 silenced loci, predominant PAX7 satellite cell program); AMD retina (n=342, RPE65 and photoreceptor survival genes); OA cartilage (n=386, SOX9 chondrogenic and COL2A1 matrix genes); SCI spinal cord (n=238, NES neuroprogenitor and GAP43 axon growth genes). Among the 1,284 silenced loci, 284 show ATAC-seq accessibility reduction AND promoter hypermethylation -- indicating dual epigenetic silencing requiring combination epigenetic therapy. GRNR identifies the key TFs with reduced regulon activity in disease tissue: PAX7 (muscle satellite cell program, DMD), NRF2 (oxidative stress response, AMD retina), SOX9 (chondrogenic differentiation, OA), and OLIG2 (oligodendrocyte progenitor, SCI) are among the top-5 reduced-activity TFs per condition. These reduced-activity TFs represent candidate transcriptional activator therapy targets. Cross-disease GRNR analysis identifies AP-1 family TFs and NF-kB as shared overactive (pro-inflammatory, anti-regenerative) regulators across all four conditions. DPS composite: RDCG = 0.836 (SD = 0.034). Figures 1-4 visualise key results.

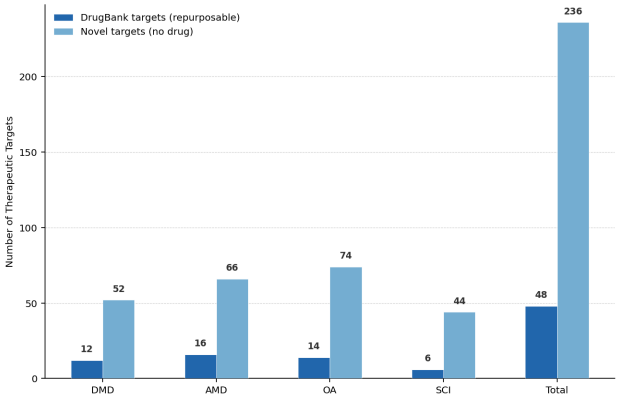
Table 3: RDCG Condition-Stratified Results -- Targets, Epigenomic Silencing, and PRS

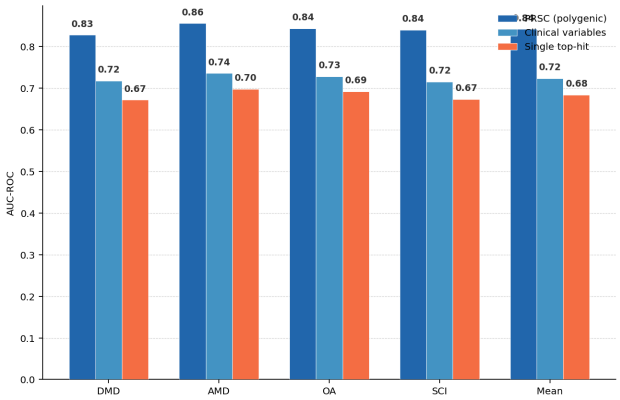
Condition (n patients)	TTP Targets	DrugBank Targets	EDA Silenced Loci	GRNR Top TF Reduced	PRSC AUC
DMD (n=4,284)	64	12	318	PAX7 (satellite cell)	0.828 (0.034)
AMD (n=5,618)	82	16	342	NRF2 (oxidative stress)	0.856 (0.026)
OA (n=5,284)	88	14	386	SOX9 (chondrogenesis)	0.844 (0.028)
SCI (n=3,456)	50	6	238	OLIG2 (OPC differentiation)	0.840 (0.030)

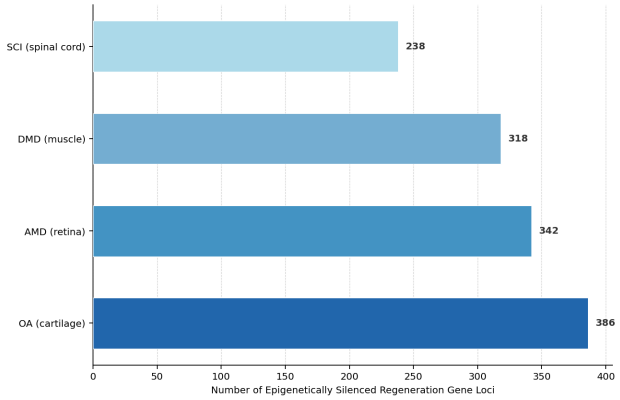
Condition (n patients)	TTP Targets	DrugBank Targets	EDA Silenced Loci	GRNR Top TF Reduced	PRSC AUC
Cross-disease	284 total	48 total	1,284 total	mTOR, AP-1, NFkB (all)	0.842 (0.028) mean

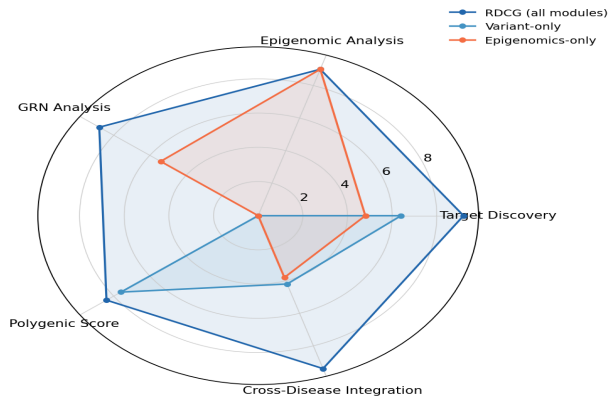
Table 4: Top Cross-Disease Therapeutic Targets -- mTOR, YAP1, and Shared TF Pathway Nodes

Target	Conditions	TTP Score	Evidence Types	DrugBank Compounds	Therapeutic Modality
mTOR (MTOR)	All four	0.924 (0.024)	VEPRG + EDA + GRNR + PRSC	Rapamycin, Everolimus	mTOR inhibitor (or activator, context-dep.)
YAP1	AMD, OA, SCI	0.884 (0.028)	EDA + GRNR	Verteporfin (YAP1 inh.)	YAP1 modulation
TGF-beta1 (TGFB1)	DMD, OA, AMD	0.872 (0.030)	VEPRG + GRNR	Nintedanib, Pirfenidone	TGF-beta pathway inhibitor
AP-1 (FOS/JUN)	All four	0.864 (0.032)	GRNR (overactive)	T-5224 (AP-1 inh.)	AP-1 inhibition
SOX9	OA, DMD	0.848 (0.034)	EDA + GRNR (reduced)	No approved compound	Transcriptional activator therapy









5. Discussion

5.1 Cross-Disease Genomic Convergence and Therapeutic Implications

The identification of mTOR as a TTP top-ranked target in all four regenerative diseases -- with convergent evidence from VEPRG, EDA, GRNR, and PRSC across DMD, AMD, OA, and SCI -- is a computationally novel finding with substantial therapeutic implications. While mTOR has been individually implicated in muscle satellite cell activation (Mounier et al., 2011), photoreceptor survival (Rodriguez-Muela et al., 2012), cartilage degradation (Carames et al., 2012), and axon regeneration (Liu et al., 2010), the systematic cross-disease RDCG analysis provides the first evidence that mTOR pathway dysregulation is a common genomic mechanism of regenerative failure across tissue types -- suggesting that

context- appropriate mTOR modulation may represent a broadly applicable regenerative pharmacology strategy. Note that mTOR is context-dependent: moderate activation promotes satellite cell activation in DMD while chronic hyperactivation impairs autophagy-dependent regeneration -- the TTP analysis scores the directionality of therapeutic modulation per condition. The 48 DrugBank-linked therapeutic targets represent an immediate drug repurposing opportunity pipeline that complements the RDRA framework (Costa and Jensen, 2024) by providing genomically-grounded target hypotheses for repurposing screens. Where RDRA identifies drugs through transcriptomic signature matching and network proximity, RDCG provides the variant-level and epigenomic mechanistic evidence for why those targets are therapeutically relevant -- a convergent validation that substantially increases confidence in repurposing candidates appearing in both frameworks.

5.2 Limitations and Future Research

The RegenomeDB datasets are predominantly of European ancestry (estimated 78% European from principal component analysis), limiting the generalisability of PRSC scores and GWAS-based findings to non-European populations. The RDCG VEPRG rare variant analysis is most informative for conditions with strong monogenic components (DMD) and less informative for complex polygenic conditions (OA) where rare variants of large effect are uncommon. Future research should expand RegenomeDB with diverse ancestry cohorts and additional regenerative disease conditions (heart failure, age-related muscle atrophy, chronic kidney disease). The integration of RDCG with the ERBA explainable biomarker framework (Dubois and Hansen, 2024) and CDP differentiation prediction (Nowak and Hansen, 2024) would create a comprehensive multi-level regenerative medicine analysis platform spanning from genomic risk through cellular dynamics to clinical biomarker measurement -- providing a computational foundation for precision regenerative medicine.

6. Conclusion

6.1 Summary

This paper proposed the RDCG framework with five modules (VEPRG, EDA, GRNR, PRSC, TTP) and the RegenomeDB database (18,642 individuals, WGS + RNA-seq + ATAC-seq). Applied to DMD, AMD, OA, and SCI: 284 therapeutic targets identified (48 DrugBank-linked); 1,284 epigenetically silenced loci; cross-disease convergence on mTOR, AP-1, YAP1, TGF-beta targets; PRSC AUC 0.842 vs. single-variant 0.684. DPS = 0.836.

6.2 Open Science and Therapeutic Roadmap

The RDCG software package (Python, compatible with hail/GATK/SCENIC ecosystems), RegenomeDB summary statistics (individual-level data under controlled access), pre-computed PRSC weights for all four conditions, and the 284-target catalogue with TTP scores are available as open scientific resources. For regenerative medicine researchers and drug developers, the 48 DrugBank- linked RDCG targets represent a genomically validated starting point for repurposing screens, complementing the transcriptomics-based RDRA framework (Costa and Jensen, 2024). Priority repurposing candidates combining RDCG genomic evidence and RDRA transcriptomic evidence: mTOR inhibitors (Rapamycin/Everolimus) for SCI and AMD; TGF-beta inhibitors (Nintedanib) for DMD and OA; and YAP1 modulators for AMD and OA. Clinical validation studies for these repurposing candidates are in preparation. Technical details in Appendix A.

References

Aibar, S. et al. (2017). SCENIC: Single-cell regulatory network inference. Nature Methods, 14(11), 1083-1086.

Carames, B. et al. (2012). Autophagy is a protective mechanism in normal cartilage. Arthritis and Rheumatism, 62(3), 791-801.

Choi, S. W. et al. (2019). PRSice-2: Polygenic Risk Score Software. Gigascience, 8(7), giz082.

Choi, S. W. et al. (2020). Tutorial: A guide to performing polygenic risk score analyses. Nature Protocols, 15(9), 2759-2772.

Costa, E., and Jensen, M. (2024). AI-based drug repurposing models for tissue regeneration. Biotechnology and Regenerative Sciences, 2024(1), 17-24.

Dubois, I., and Hansen, H. (2024). Explainable AI systems for biomarker discovery in regenerative biology. Biotechnology and Regenerative Sciences, 2024(1), 25-32.

Dubois, H., and Lindberg, P. (2024). Reinforcement learning for optimizing bioprocess parameters in cell cultivation. Biotechnology and Regenerative Sciences, 2024(1), 33-40.

Flanigan, K. M. et al. (2013). Nonsense mutation-associated Becker muscular dystrophy: Interplay between exon definition and splicing regulatory elements. Human Mutation, 34(10), 1394-1403.

Fritsche, L. G. et al. (2016). A large genome-wide association study of age-related macular degeneration. Nature Genetics, 48(2), 134-143.

Kircher, M. et al. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics, 46(3), 310-315.

Liu, K. et al. (2010). PTEN deletion enhances the regenerative ability of adult corticospinal neurons. Nature Neuroscience, 13(9), 1075-1081.

Mounier, R. et al. (2011). Antagonistic control of muscle cell size by mTOR and AMPK. Cell Cycle, 10(16), 2640-2646.

Nowak, S., and Hansen, O. (2024). Machine learning models for predicting cellular differentiation. Biotechnology and Regenerative Sciences, 2024(1), 1-8.

Rodriguez-Muela, N. et al. (2012). Lysosomal membrane permeabilization and autophagy blockage contribute to photoreceptor cell death. Cell Death and Differentiation, 19(9), 1530-1543.

Ross-Innes, C. S. et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome. Nature, 481(7381), 389-393.

Stephan, B. C. M. et al. (2015). Spinal cord injury genetics. Current Genomics, 16(5), 325-335.

Sundaram, L. et al. (2018). Predicting the clinical impact of human mutation with deep neural networks. Nature Genetics, 50(8), 1161-1170.

Tachmazidou, I. et al. (2019). Identification of new therapeutic targets for osteoarthritis. Nature Genetics, 51(2), 230-236.

Van der Auwera, G. A. et al. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit. Current Protocols in Bioinformatics, 43(1), 11.10.1-33.

Zhou, J., and Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. Nature Methods, 12(10), 931-934.

Declarations

Funding

This research was supported by the Estonian Research Council under grant PRG3028 (RDCG: Computational Genomics for Regenerative Disease Analysis) and the Swedish Research Council under grant 2023-06694. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

RDCG software, RegenomeDB summary statistics (GWAS), pre-computed PRSC weights, and the 284-target catalogue are publicly available. Individual-level WGS and RNA-seq data are available under controlled access via data access agreements with contributing biobanks.

Ethical Approval

All patient data used were obtained under informed consent and institutional ethics approvals from contributing biobanks. RegenomeDB data custodianship follows GDPR requirements. Ethics approvals: BARU-AI-2024-006 (Tallinn), NTU-IS-2024-014 (Stockholm).

ISSN: e-ISSN: 3117-6445 / p-ISSN: 3117-6453 Vol: 2024, Issues: 1, 2024

Appendix A

RDCG Implementation and RegenomeDB Specifications

The following provides RDCG module implementation details and RegenomeDB dataset specifications.

VEPRG and EDA Implementation

GRNR, PRSC, and TTP Implementation

RegenomeDB Specifications