aiAtlas: High-Fidelity Cell Simulations of Genetic Perturbations in Rare Diseases and Cancers

Wayne R Danter, MD

Abstract

Background: Understanding how genetic perturbations drive cellular phenotypes remains a central challenge in disease modeling. While induced pluripotent stem cells (iPSCs) and tumor-derived cell lines have advanced mechanistic research, they are constrained by scalability, stability, and limited capacity to reproduce complex mutational states.

Methods: We developed aiAtlas, a reliable computational system that integrates Large Concept Model (LCM) based causal inference with aiPSC derived simulation libraries. The platform encodes 25 core biological and cellular features, including apoptosis, DNA damage and repair pathways, cell cycle regulation, pluripotency markers, and the hallmarks of cancer, in weighted causal networks. Simulations were iteratively propagated using early-stopping algorithms, with statistical evaluation performed using Mann Whitney U tests, Hodges Lehmann estimators (HLE), and Cliff's delta with Hedges correction. Bonferroni-adjusted significance thresholds ($p < 0.002$) to compensate for multiple testing were applied.

Results: Across 136 simulated aiPSC lines (10 wild type, 126 mutants), all 25 features demonstrated significant divergence between wild type and combined mutant states (median HLE -1.46, IQR -1.89 to -0.36). Subgroup analyses revealed distinct patterns: single mutations drove broad systemic changes; multiple mutations amplified genomic instability; human tumor-derived cell lines exhibited strong divergence but retained some overlap in repair and apoptosis pathways; and gene fusion models produced selective, intermediate phenotypes. Effect sizes were consistently large (Cliff's delta approaching -0.95 with 95% CIs).

Conclusions: aiAtlas reliably distinguishes wild type from mutant cell states, capturing both global and subgroup-specific mutational divergence. By combining causal inference with LCM-level abstraction, aiAtlas establishes a scalable, reproducible, and physiologically grounded platform for rare disease modeling, cancer biology, and therapeutic discovery.

Significance Statement

aiAtlas introduces a scalable, reliable simulation framework that integrates advanced Large Concept Model (LCM) logic with iPSC-derived cellular modeling. aiAtlas overcomes major limitations of experimental systems by capturing both broad and subgroup-specific phenotypic divergence across single mutations, multiple mutations, tumor-derived cell lines, and gene fusions. This reliable platform establishes a new opportunity for rare

diseases and cancer modeling, offering reproducible insights that can accelerate discovery and translational applications where traditional wet-lab approaches are impractical.

Furthermore, in situations where the target mutational profile has been defined but no cellular models yet exist, aiAtlas can quickly generate custom virtual cell lines that precisely reproduce the corresponding genomic and phenotypic features.

Keywords: aiPSC cell lines; rare and ultra-rare diseases; large concept model (LCM); DNA damage and repair; epigenetic reprogramming; cancer modeling; computational biology; precision medicine.

Introduction

Understanding how genetic alterations drive disease phenotypes at the cellular level remains one of the fundamental challenges in biomedical research. While induced pluripotent stem cells (iPSCs) and established human cell lines have provided powerful experimental systems, they remain constrained by cost, scalability, and an inability to fully capture the systemic consequences of complex genomic states [1,2]. These limitations are particularly important in the context of rare and ultra-rare diseases, where patient material is often scarce [3], and in cancers, where multiple mutations and gene fusions often interact in unpredictable ways [4,5].

While evolving computational models attempt to bridge this gap, most current approaches still lack the fidelity to reproduce multi-pathway interactions at cellular and systems levels [6]. There remains an unmet need for a platform that can integrate genomic, epigenomic, and cellular features to produce more physiologically realistic, and testable predictions.

We recently developed aiAtlas, a simulation-based platform built on large concept model (LCM) logic and aiPSC modeling, designed to simulate cellular states and predict phenotypic divergence arising from a broad array of genetic perturbations [6,7]. Large Concept Models (LCMs) represent a significant advancement over traditional Fuzzy Cognitive Maps (FCMs). While FCMs are constrained to explicit networks of weighted causal relationships defined between user-specified concepts (nodes), LCMs operate in a high-dimensional concept space that allows for abstraction, generalization, and the emergence of new properties beyond predefined variables. LCMs represent a powerful AI architecture that extends beyond basic causal linkages to capture the broader conceptual dynamics of complex systems, making them a powerful evolution of FCM logic for high-fidelity biological simulation.

In Part 1, we evaluate the ability of aiAtlas to reliably distinguish wild type (WT) aiPSCs from a diverse population of mutant cells, capturing major shifts in cell cycle regulation, DNA repair, epigenetic remodeling, pluripotency, and the hallmarks of cancers.

In Part 2 of this study, we extend this framework to clinically relevant subgroups, each representing a distinct biological context including single gene mutations, multiple mutations, human tumor-derived cell lines, and gene fusions. By systematically comparing these four subgroups against WT aiPSCs, we aim to determine whether aiAtlas can not only detect broad divergence but also resolve more granular, subgroup-specific phenotypes. The current study aims to evaluate aiAtlas as a scalable, high-fidelity artificial intelligence (AI) platform capable of simulating pluripotency states, rare diseases, and cancer genetics.

Methods

aiAtlas Simulation Platform

All analyses were performed using aiAtlas, a high-resolution computational modeling framework that integrates large concept model (LCM)-based causal inference with aiPSC simulation libraries. The platform simulates cellular states by combining genetic, epigenetic, and signaling features into interconnected causal networks. Each concept (node) represents a measurable biological process, and edges define directional and weighted causal relationships between nodes on a continuous -1 to +1 scale. Simulations were iteratively propagated until optimal early stopping states were reached. Early stopping was used as a regularization method to improve generalizability and reduce the potential for overfitting the data.

Cell Line Cohorts

In Part 1, we analyzed 136 aiPSC lines using the aiAtlas platform, comparing 10 wild type (WT) lines with 126 mutant lines. In Part2, the simulated aiPSC lines were grouped into four experimental cohorts for comparison against WT (N = 10): single mutations (N = 81), multiple mutations (N = 20), human tumor-derived cell lines (N = 10), and gene fusion lines (N = 15). All simulations were run under identical baseline conditions, with differences arising only from input mutational profiles. All aiPSC wild type (WT) and mutated lines were generated using the aiCRISPR-Like gene editing simulation technology [8].

Feature Definitions

Twenty-five biological and output features were evaluated and grouped into categories: cell cycle regulation, apoptosis and autophagy, DNA damage and repair, epigenetic remodeling, pluripotency/self-renewal, oncogenic features, stress responses, etc.). Each feature was scaled to the [-1, +1] range, where negative values represent reduced or dysregulated function relative to the WT cellular state. Supporting references for all selected features and definitions are listed in Appendix A.

Statistical Analysis, etc.

Pairwise group comparisons were performed between WT and each mutated group or subgroup. The non-parametric Mann-Whitney U test was used to assess distributional differences [9]. An adjusted significance threshold with the conservative Bonferroni correction for multiple testing [10] was applied ($p < 0.002$). The Hodges-Lehmann Estimator (HLE) was reported as the median of all pairwise differences between groups [11], presented as a point estimate without confidence intervals. Cliff's delta with Hedges-G correction [12,13] was reported as a nonparametric effect size, with 95% confidence intervals.

Validation and Reproducibility

All simulations were repeated independently to confirm stability of outcomes. Internal quality control included convergence monitoring, 5-fold cross-validation, bootstrap resampling, system wide error, and stability indices. Together, these measures ensured reproducibility and robustness of aiAtlas outputs across cohorts.

Results

We analyzed 136 aiPSC lines using the aiAtlas platform, comparing 10 wild type (WT) lines with 126 mutant lines. 25 biological and cellular features were evaluated using nonparametric methods. Group differences were tested with the Mann-Whitney U test. Effect sizes were summarized in two forms: the Hodges-Lehmann Estimator (HLE), representing the median of all pairwise differences between groups, and Cliff's delta with Hedges-G correction, included as a dedicated feature entry in the dataset. 95% confidence intervals (CIs) were computed for Cliff's delta effect sizes values.

In Part 1 we compared all 126 mutated aiPSC derived cell lines to 10 Wild Type aiPSC. In Part 2 we divided the data to create 4 specific subgroups for comparison with the aiPSC Wild Type virtual cells.

Part1: Overall Findings

All 25 features demonstrated statistically significant differences (corrected $p < 0.002$) between WT and mutant aiPSC lines. The median HLE across all features was -1.46 (IQR -1.89 to -0.36). Several features approached maximum values (HLE ~ -1.9), indicating nearly complete separation of WT and mutant cell distributions. Cliff's delta effect size estimates with Hedges correction, confirmed consistently large effect sizes, (e.g., Cliffs Delta with Hedges-G Correction = -0.926, $p = 1.57 \times 10^{-6}$).

Key Representative Features

Table 1:

Feature	HLE	Mann-Whitney u	Cliff's delta	95%CI ±
Apoptosis_Extrinsic pathway	-0.100	1.65E-06	-0.869	0.306
Apoptosis_Intrinsic pathway	-0.132	2.75E-05	-0.756	0.406
DNA_Damage_Burden/Accumulation	-0.335	1.53E-07	-0.952	0.189
DNA_Damage_CPD/6-4_PPs	-0.585	8.30E-12	-0.922	0.240
G1S_Transition	-1.955	9.22E-08	-0.952	0.189
G2M_Transition	-1.470	1.53E-07	-0.952	0.189
Genome_Instability/CIN	-0.447	1.35E-06	-0.877	0.298
Marker_Stability	-1.530	1.53E-07	-0.952	0.189
The Hallmarks of Cancer	-1.890	1.48E-07	-0.952	0.189

Part 1: Interpretation

The Hodges-Lehmann Estimator (HLE) values confirm variable but substantial distributional shifts between WT and mutant lines, particularly in cell cycle regulation, oncogenic hallmark pathways, and overall marker stability. Moderate but significant shifts were also observed in DNA damage repair and apoptosis. The summary Cliff's delta with Hedges correction confirms independently verified these large effect sizes. Together, these results support robust divergence of mutant aiPSCs from WT baselines across multiple biological dimensions. Full results for all 25 features, including HLE values, Mann-Whitney p-values, and the Cliff's delta estimates with 95% CIs, are provided in Supplementary Table S1a.

Part 2: Subgroup Comparisons

WT vs Single Mutations (single aiCRISPR-L gene edits)

Single mutations (N = 81) produced broad systemic disruption, with nearly complete separation from WT across cell cycle, DNA repair, epigenetic, and pluripotency features. Cliff's delta values approached -0.95 for most features, confirming very large effect sizes. In general, HLE median differences were more variable. Full results for all 25 features, including HLE values, Mann-Whitney p-values, and the Cliff's delta estimates with 95% CIs, are provided in Supplementary Table S2a.

Table 2: aiPSC WT vs Single Mutations

Feature	HLE	Mann-Whitney u	Cliff's delta	95%CI ±
Apoptosis_Extrinsic pathway	-0.100	2.81E-07	-0.952	0.189
Apoptosis_Intrinsic pathway	-0.132	2.86E-11	-0.929	0.230
DNA_Damage_Burden/Accumulation	-0.291	2.82E-07	-0.952	0.189
DNA_Damage_CPD/6-4_PPs	-0.560	2.82E-07	-0.952	0.189
G1S_Transition	-1.958	1.24E-07	-0.952	0.189
G2M_Transition	-1.545	2.82E-07	-0.952	0.189
Genome_Instability/CIN	-0.436	2.82E-07	-0.952	0.189
Marker_Stability	-1.534	2.82E-07	-0.952	0.189
The Hallmarks of Cancer	-1.892	2.67E-07	-0.952	0.189

WT vs Multiple Mutations (multiple aiCRISPR-L gene edits)

Multiple mutations (N = 20) amplified instability, mirroring clinical disease progression. While most features showed large effects, intrinsic apoptosis and NER-Core exhibited more moderate differences, reflecting biological heterogeneity. Again, HLE median differences were more variable. Full results for all 25 features, including HLE values, Mann-Whitney p-values, and the Cliff's delta estimates with 95% CIs, are provided in Supplementary Table S3a.

Table 3: aiPSC WT vs Multiple Mutations

Feature	HLE	Mann-Whitney u	Cliff's delta	95%CI ±
Apoptosis_Extrinsic pathway	-0.100	6.12E-06	-0.857	0.319
Apoptosis_Intrinsic pathway	-0.111	2.50E-02	-0.476	0.545
DNA_Damage_Burden/Accumulation	-0.455	9.90E-06	-0.952	0.189
DNA_Damage_CPD/6-4_PPs	-0.567	9.90E-06	-0.952	0.189
G1S_Transition	-1.959	7.98E-06	-0.952	0.189

G2M_Transition	-1.538	9.90E-06	-0.952	0.189
Genome_Instability/CIN	-0.548	9.90E-06	-0.952	0.189
Marker_Stability	-1.618	9.87E-06	-0.952	0.189
The Hallmarks of Cancer	-1.910	9.69E-06	-0.952	0.189

WT vs Human Cell Lines (multiple aiCRISPR-L gene edits to create the actual cell line profile)

Widely available human tumor-derived cell lines (N = 10) diverged strongly from WT in cell cycle regulation, pluripotency, and genomic stability. However, partial overlap was observed in intrinsic apoptosis and repair pathways, consistent with the adaptive survival mechanisms of immortalized lines. In a few cases, HLE median differences were also more variable. Full results for all 25 features, including HLE values, Mann-Whitney p-values, and the Cliff's delta estimates with 95% CIs, are provided in Supplementary Table S4a.

Table 4: aiPSC WT vs Human Cell Lines

Feature	HLE	Mann-Whitney u	Cliff's delta	95%CI ±
Apoptosis_Extrinsic pathway	-0.096	9.96E-04	-0.762	0.401
Apoptosis_Intrinsic pathway	0.006	9.86E-01	0.000	0.620
DNA_Damage_Burden/Accumulation	-0.600	1.10E-04	-0.952	0.189
DNA_Damage_CPD/6-4_PPs	-0.613	1.79E-02	-0.571	0.509
G1S_Transition	-1.961	1.07E-04	-0.952	0.189
G2M_Transition	-1.751	1.10E-04	-0.952	0.189
Genome_Instability/CIN	-1.591	1.09E-04	-0.952	0.189
Marker_Stability	-1.753	1.10E-04	-0.952	0.189
The Hallmarks of Cancer	-1.932	1.10E-04	-0.952	0.189

WT vs Gene Fusion Cells (specific gene fusions via aiCRISPR-L gene edits)

aiCRISPR-Like edited gene fusion models (N = 15) displayed intermediate phenotypes. Strong differences were observed in cell cycle, DNA repair, and pluripotency markers, but apoptosis and genome instability showed weaker or nonsignificant effects. Fusion-driven phenotypes were thus more selective compared to multiple mutations. Full results for all 25

features, including HLE values, Mann-Whitney p-values, and the Cliff's delta estimates with 95% CIs, are provided in Supplementary Table S5a.

Table 5: aiPSC WT vs Gene Fusions

Feature	HLE	Mann-Whitney u	Cliff's delta	95%CI ±
Apoptosis_Extrinsic pathway	0.000	2.04E-02	-0.508	0.534
Apoptosis_Intrinsic pathway	-0.057	1.10E-03	-0.698	0.444
DNA_Damage_Burden/Accumulation	-0.296	2.65E-05	-0.952	0.189
DNA_Damage_CPD/6-4_PPs	-0.786	2.65E-05	-0.952	0.189
G1S_Transition	-0.958	2.65E-05	-0.952	0.189
G2M_Transition	-0.914	2.65E-05	-0.952	0.189
Genome_Instability/CIN	-0.061	1.66E-01	-0.317	0.588
Marker_Stability	-1.117	2.65E-05	-0.952	0.189
The Hallmarks of Cancer	-0.836	2.64E-05	-0.952	0.189

Part 2: Integrated Interpretation

Across all comparisons, aiAtlas consistently detected significant divergence from WT baselines. Single and multiple mutations produced system-wide disruption, human cell lines showed tumor-like instability with partial functional retention, and gene fusions induced selective and less global divergence. Together, these analyses validate aiAtlas as a robust platform for modeling rare disease-relevant mutations, complex genomic states, and cancer-associated rearrangements.

Discussion

This study demonstrates that aiAtlas provides a high-fidelity computational framework for modeling genetic perturbations across diverse biological contexts in virtual cell lines. Part 1 established broad divergence between WT and diverse mutant lines, while Part 2 revealed subgroup-specific patterns. Single mutations were sufficient to produce systemic divergence, multiple mutations amplified instability, human tumor cell lines diverged but retained partial overlap with WT, and gene fusions induced intermediate, more selective phenotypes.

These results highlight both shared mechanisms (cell cycle dysregulation, pluripotency loss) and subgroup-specific nuances (variability in apoptosis and DNA repair). By capturing both global and granular phenotypes, aiAtlas bridges the gap between reductionist models and the complex, multifactorial nature of human disease at the cellular level. Prior studies have shown that iPSCs, while invaluable, remain limited by scalability, stability, and patient sample scarcity in rare diseases [1-3], and by the difficulty of reproducing complex mutational interactions in cancers [4,5]. Our findings align with and extend these observations by demonstrating that aiAtlas, through integration of Large Concept Model (LCM) logic [6], can systematically replicate such divergences in silico.

The implications of this study are important for rare disease research, cancer biology, and therapeutic discovery. Patient-derived iPSCs have shown promise for modeling rare genetic disorders but face bottlenecks in cost and accessibility [3]. Similarly, modeling cancer in vitro with iPSCs or immortalized cell lines has highlighted the challenges of multiple cooperating mutations and fusion-driven oncogenesis [4,5]. Computational models have attempted to bridge these gaps, but existing approaches often fail to reproduce multi-pathway interactions [7]. By combining FCM-based causal inference with LCM abstraction, aiAtlas provides a scalable and reproducible alternative that captures both high-level and mechanistic cellular dynamics.

Limitations include the need for continued benchmarking against experimental iPSC and organoid datasets, and the extension of aiAtlas to incorporate metabolic, microenvironmental, and immune system features - critical drivers of disease not yet fully modeled [1,7]. Nevertheless, the current study demonstrates that aiAtlas can reliably reproduce physiologically relevant divergence across multiple genomic contexts.

In conclusion, aiAtlas establishes a new standard for computational disease modeling, providing reproducible and physiologically grounded predictions. This platform offers potential to accelerate discovery and improve translational outcomes in rare diseases and oncology, consistent with the broader need for innovative, scalable systems that overcome the limitations of traditional wet-lab approaches [1-3,7].

Importantly, in cases where the target mutational profile has been defined but no cellular models yet exist, aiAtlas can quickly generate custom virtual cell lines that precisely reproduce the corresponding genomic and phenotypic features.

Reference List

1. Cernickis P, et al. iPSC technology in 2024: advances, limitations, and future perspectives. Signal Transduct Target Ther. 2024;9:221. doi:10.1038/s41392-024-01617-8.
2. Moy Y, et al. iPSC-based therapies: translational challenges and opportunities. Front Cell Dev Biol. 2023;11:10768945. doi:10.3389/fcell.2023.10768945.
3. Francis N, et al. Patient-derived iPSCs for rare genetic disorders: promise and bottlenecks. Front Cell Dev Biol. 2025;13:1442039. doi:10.3389/fcell.2025.1442039.

4. Kim J, et al. Modeling cancer using patient-derived iPSCs: opportunities and challenges. Front Immunol. 2018;9:188. doi:10.3389/fimmu.2018.00188.
5. Stransky N, et al. The landscape of kinase fusions in cancer. Nat Commun. 2014;5:4846. doi:10.1038/ncomms5846.
6. Barrault L, et al. Large Concept Models: Language Modeling in a Sentence Representation Space. arXiv preprint arXiv:2412.08821. 2024.
7. Shmulevich I, et al. Challenges and opportunities in systems biology modeling. Cell Syst. 2022;13(8):611-627. doi:10.1016/j.cels.2022.07.005.
8. Danter WR. aiCRISPRL: An artificial intelligence platform for stem cell and organoid simulation with extensive gene editing capabilities. bioRxiv. 2024 Jun 13. doi:10.1101/2022.06.18.496679v2.
9. Nachar N. The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. Tutor Quant Methods Psychol. 2008;4(1):13-20. doi:10.20982/tqmp.04.1.p013.
10. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ. 1995;310(6973):170. doi:10.1136/bmj.310.6973.170.
11. Hollander M, Wolfe DA, Chicken E. Nonparametric Statistical Methods. 3rd ed. Hoboken (NJ): Wiley; 2013.
12. Cliff N. Dominance statistics: Ordinal analyses to answer ordinal questions. Psychol Bull. 1993;114(3):494-509. doi:10.1037/0033-2909.114.3.494.
13. Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. San Diego: Academic Press; 1985.

Table S1a:
aiPSC Wild Type vs All Mutated Lines

Concept	HLE	Mann-Whitney u	Cliff's delta	95%CI
Apoptosis_Extrinsic pathway	-0.100	1.65E-06	-0.869	0.306
Apoptosis_Intrinsic pathway	-0.132	2.75E-05	-0.756	0.406
Autophagy	-0.407	1.53E-07	-0.952	0.189
Cellular stress	-0.051	1.51E-07	-0.952	0.189
DNA_Damage_Burden/Accumulation	-0.335	1.53E-07	-0.952	0.189
DNA_Damage_CPD/6-4_PPs	-0.585	8.30E-12	-0.922	0.240
DNA_NER_Core	-0.824	2.15E-05	-0.771	0.395
DNA_NER-GG	-0.078	1.53E-07	-0.952	0.189

DNA_NER-TC	-0.156	1.53E-07	-0.952	0.189
DNA_Replication Stress	-0.364	1.53E-07	-0.952	0.189
Epigenetic Dysfunction	-0.353	1.53E-07	-0.952	0.189
ERCC2_Function/Efficiency	-0.079	5.51E-10	-0.890	0.282
G1S_Transition	-1.955	9.22E-08	-0.952	0.189
G2M_Transition	-1.470	1.53E-07	-0.952	0.189
Genome_Instability/CIN	-0.447	1.35E-06	-0.877	0.298
gH2AX	-1.157	1.53E-07	-0.952	0.189
HypermethylationDNA	-1.943	1.52E-07	-0.952	0.189
Hypomethylation/DemethylationDNA	-0.806	1.53E-07	-0.952	0.189
Marker_Stability	-1.530	1.53E-07	-0.952	0.189
Mitochondrial_Stress	-0.026	8.17E-12	-0.922	0.240
Nanog	-1.961	8.50E-08	-0.952	0.189
OCT_3/4	-1.908	8.60E-08	-0.952	0.189
ROS/OxidativeStress	-0.029	2.44E-08	-0.952	0.189
Sox2	-1.967	3.18E-08	-0.952	0.189
The Hallmarks of Cancer	-1.890	1.48E-07	-0.952	0.189

Table S2a
WT vs Single Mutations

Feature	HLE	Mann-Whitney u	Cliff's delta	95%CI
[007_Apoptosis_Extrinsic pathway]	-0.100	2.81E-07	-0.952	0.189
[012_Autophagy]	-0.381	2.82E-07	-0.952	0.189

[017_G1S_Transition]	-1.958	1.24E-07	-0.952	0.189
[018_G2M_Transition]	-1.545	2.82E-07	-0.952	0.189
[021_Apoptosis_Intrinsic pathway]	-0.132	2.86E-11	-0.929	0.230
[037_The Hallmarks of Cancer	-1.892	2.67E-07	-0.952	0.189
Cellular stress	-0.052	2.77E-07	-0.952	0.189
DNA_Damage_Burden/Accumulation	-0.291	2.82E-07	-0.952	0.189
DNA_Damage_CPD/6-4_PPs	-0.560	2.82E-07	-0.952	0.189
DNA_NER_Core	-0.861	5.2E-09	-0.882	0.292
DNA_NER-GG	-0.075	2.82E-07	-0.952	0.189
DNA_NER-TC	-0.152	2.82E-07	-0.952	0.189
DNA_Replication Stress	-0.369	2.82E-07	-0.952	0.189
Epigenetic Dysfunction	-0.355	2.82E-07	-0.952	0.189
ERCC2_Function/Efficiency	-0.076	2.82E-07	-0.952	0.189
Genome_Instability/CIN	-0.436	2.82E-07	-0.952	0.189
gH2AX	-1.237	2.82E-07	-0.952	0.189
HypermethylationDNA	-1.954	2.81E-07	-0.952	0.189
Hypomethylation/DemethylationDNA	-0.837	2.82E-07	-0.952	0.189
Marker_Stability	-1.534	2.82E-07	-0.952	0.189
Mitochondrial_Stress	-0.026	2.81E-07	-0.952	0.189
Nanog	-1.963	8.81E-08	-0.952	0.189
OCT_3/4	-1.965	1.33E-07	-0.952	0.189
ROS/OxidativeStress	-0.029	5.97E-09	-0.952	0.189
Sox2	-1.967	4.97E-08	-0.952	0.189

Table S3a
WT vs Multiple Mutations

Feature	HLE	Mann-Whitney u	Cliff's delta	95%CI
[007_Apoptosis_Extrinsic pathway]	-0.100	6.12E-06	-0.857	0.319
[012_Autophagy]	-0.411	9.90E-06	-0.952	0.189
[017_G1S_Transition]	-1.959	7.98E-06	-0.952	0.189
[018_G2M_Transition]	-1.538	9.90E-06	-0.952	0.189
[021_Apoptosis_Intrinsic pathway]	-0.111	2.50E-02	-0.476	0.545
[037_The Hallmarks of Cancer	-1.910	9.69E-06	-0.952	0.189
Cellular stress	-0.059	9.90E-06	-0.952	0.189
DNA_Damage_Burden/Accumulation	-0.455	9.90E-06	-0.952	0.189
DNA_Damage_CPD/6-4_PPs	-0.567	9.90E-06	-0.952	0.189
DNA_NER_Core	-0.459	6.16E-03	-0.571	0.509
DNA_NER-GG	-0.035	9.81E-06	-0.952	0.189
DNA_NER-TC	-0.048	9.87E-06	-0.952	0.189
DNA_Replication Stress	-0.384	9.90E-06	-0.952	0.189
Epigenetic Dysfunction	-0.378	9.90E-06	-0.952	0.189
ERCC2_Function/Efficiency	-0.062	2.00E-07	-0.943	0.207
Genome_Instability/CIN	-0.548	9.90E-06	-0.952	0.189
gH2AX	-1.291	9.90E-06	-0.952	0.189
HypermethylationDNA	-1.943	9.67E-06	-0.952	0.189
Hypomethylation/DemethylationDNA	-0.940	9.90E-06	-0.952	0.189
Marker_Stability	-1.618	9.87E-06	-0.952	0.189
Mitochondrial_Stress	-0.031	9.90E-06	-0.952	0.189

Nanog	-1.967	7.98E-06	-0.952	0.189
OCT_3/4	-1.967	6.59E-06	-0.952	0.189
ROS/OxidativeStress	-0.029	6.59E-06	-0.952	0.189
Sox2	-1.967	3.06E-06	-0.952	0.189

Table S4a
WT vs Human Cell Lines

Feature	HLE	Mann-Whitney u	Cliff's delta	95%CI
[007_Apoptosis_Extrinsic pathway]	-0.096	9.96E-04	-0.762	0.401
[012_Autophagy]	-0.201	1.10E-04	-0.952	0.189
[017_G1S_Transition]	-1.961	1.07E-04	-0.952	0.189
[018_G2M_Transition]	-1.751	1.10E-04	-0.952	0.189
[021_Apoptosis_Intrinsic pathway]	0.006	9.86E-01	0.000	0.620
[037_The Hallmarks of Cancer	-1.932	1.10E-04	-0.952	0.189
Cellular stress	-0.157	1.10E-04	-0.952	0.189
DNA_Damage_Burden/Accumulation	-0.600	1.10E-04	-0.952	0.189
DNA_Damage_CPD/6-4_PPs	-0.613	1.79E-02	-0.571	0.509
DNA_NER_Core	-0.001	9.87E-01	0.000	0.620
DNA_NER-GG	-0.010	1.09E-04	-0.952	0.189
DNA_NER-TC	-0.002	1.07E-04	-0.952	0.189
DNA_Replication Stress	-0.671	1.10E-04	-0.952	0.189
Epigenetic Dysfunction	-0.523	1.10E-04	-0.952	0.189
ERCC2_Function/Efficiency	-0.006	4.54E-01	-0.190	0.608

Genome_Instability/CIN	-1.591	1.09E-04	-0.952	0.189
gH2AX	-0.804	1.10E-04	-0.952	0.189
HypermethylationDNA	-1.931	1.10E-04	-0.952	0.189
Hypomethylation/DemethylationDNA	-0.861	1.10E-04	-0.952	0.189
Marker_Stability	-1.753	1.10E-04	-0.952	0.189
Mitochondrial_Stress	-0.047	1.10E-04	-0.952	0.189
Nanog	-1.967	9.66E-05	-0.952	0.189
OCT_3/4	-1.963	1.09E-04	-0.952	0.189
ROS/OxidativeStress	-0.033	1.09E-04	-0.952	0.189
Sox2	-1.967	9.66E-05	-0.952	0.189

Table S5a
WT vs Gene Fusions

Feature	HLE	Mann-Whitney u	Cliff's delta	95%CI
[007_Apoptosis_Extrinsic pathway]	0.000	2.04E-02	-0.508	0.534
[012_Autophagy]	-1.490	2.65E-05	-0.952	0.189
[017_G1S_Transition]	-0.958	2.65E-05	-0.952	0.189
[018_G2M_Transition]	-0.914	2.65E-05	-0.952	0.189
[021_Apoptosis_Intrinsic pathway]	-0.057	1.10E-03	-0.698	0.444
[037_The Hallmarks of Cancer	-0.836	2.64E-05	-0.952	0.189
Cellular stress	-0.046	2.12E-05	-0.952	0.189
DNA_Damage_Burden/Accumulation	-0.296	2.65E-05	-0.952	0.189
DNA_Damage_CPD/6-4_PPs	-0.786	2.65E-05	-0.952	0.189

DNA_NER_Core	-0.780	2.65E-05	-0.952	0.189
DNA_NER-GG	-0.564	2.65E-05	-0.952	0.189
DNA_NER-TC	-0.673	2.65E-05	-0.952	0.189
DNA_Replication Stress	-0.193	2.65E-05	-0.952	0.189
Epigenetic Dysfunction	-0.147	2.65E-05	-0.952	0.189
ERCC2_Function/Efficiency	-0.497	2.65E-05	-0.952	0.189
Genome_Instability/CIN	-0.061	1.66E-01	-0.317	0.588
gH2AX	-0.585	2.65E-05	-0.952	0.189
HypermethylationDNA	-1.130	2.65E-05	-0.952	0.189
Hypomethylation/DemethylationDNA	-0.282	2.65E-05	-0.952	0.189
Marker_Stability	-1.117	2.65E-05	-0.952	0.189
Mitochondrial_Stress	-0.004	1.06E-03	-0.698	0.444
Nanog	-0.967	2.65E-05	-0.952	0.189
OCT_3/4	-0.863	2.65E-05	-0.952	0.189
ROS/OxidativeStress	-0.028	2.51E-05	-0.952	0.189
Sox2	-1.000	2.65E-05	-0.952	0.189

Appendix A. aiPSC Cell Line Concepts and Supporting References

Apoptosis - Extrinsic Pathway

1. Ashkenazi A, Dixit VM. Death receptors: signaling and modulation. Science. 1998;281(5381):1305-1308.
2. Fulda S, Debatin KM. Apoptosis signaling in tumor therapy. Ann N Y Acad Sci. 2004;1030:150-159.
3. Lavrik IN. Systems biology of death receptor networks: live and let die. Cell Death Dis. 2014;5(5):e1259.

Apoptosis - Intrinsic Pathway

4. Youle RJ, Strasser A. The BCL-2 protein family: opposing activities that mediate cell death. Nat Rev Mol Cell Biol. 2008;9(1):47-59.

5. Czabotar PE, Lessene G, Strasser A, Adams JM. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. Nat Rev Mol Cell Biol. 2014; 15(1):49-63. doi:10.1038/nrm3722.
6. Tait SWG, Green DR. Mitochondria and cell death: outer membrane permeabilization and beyond. Nat Rev Mol Cell Biol. 2010;11(9):621-632.

Autophagy

7. Mizushima N, Levine B. Autophagy in mammalian development and differentiation. Nat Cell Biol. 2010;12(9):823-830.
8. Guan JL, Simon AK, Prescott M, et al. Autophagy in stem cells. Autophagy. 2013;9(6):830-849. doi:10.4161/auto.24132.
9. Xu Y, Zhang Y, García-Cañaveras JC, et al. Chaperone-mediated autophagy regulates the pluripotency of embryonic stem cells. Science. 2020;369(6502):397-403.

Cellular Stress

10. Mandal PK, Rossi DJ. Pluripotent stem cells and DNA damage responses. Nat Rev Mol Cell Biol. 2013;14(7):443-458.
11. Martello G, Smith A. The nature of embryonic stem cells. Annu Rev Cell Dev Biol. 2014;30:647-675.

DNA Damage Burden / Accumulation

12. Momcilovic O, Knobloch L, Fornsaglio J, Varum S, Easley C, Schatten G. DNA damage responses in human induced pluripotent stem cells and embryonic stem cells. Cell Stem Cell. 2010;7(3):329-342.
13. Li X, Qu K, Zhao C, et al. DNA repair in pluripotent stem cells. DNA Repair (Amst). 2012;11(6):589-600.
14. González F, Georgieva D, Vanoli F, et al. DNA damage and repair during reprogramming to pluripotency. Cell Stem Cell. 2013;13(4):360-369.

DNA Damage - CPDs / 6-4 PPs

15. Pfeifer GP, You YH, Besaratinia A. Mutations induced by ultraviolet light. Mutat Res. 2005;571(1-2):19-31.
16. Rastogi RP, Richa, Kumar A, Tyagi MB, Sinha RP. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. J Nucleic Acids. 2010;2010:592980.
17. Schärer OD. Nucleotide excision repair in eukaryotes. Cold Spring Harb Perspect Biol. 2013;5(10):a012609.

DNA NER - Core Pathway

18. Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. Understanding nucleotide excision repair and its roles in cancer and ageing. Nat Rev Mol Cell Biol. 2014;15(7):465-481.
19. Schärer OD. Nucleotide excision repair in eukaryotes. Cold Spring Harbor Perspectives in Biology. 2013;5(10):a012609.
20. Cleaver JE. Defective repair replication of DNA in xeroderma pigmentosum. J Invest Dermatol. 2005;124(3):xv-xix.

DNA NER - Global Genome (GG-NER)

21. Hanawalt PC. Role of global genome nucleotide excision repair in preventing cancer. Mutat Res. 2020;821:111692.
22. Riedl T, Hanaoka F, Egly JM. The comings and goings of nucleotide excision repair factors on damaged DNA. EMBO J. 2003;22(19):5293-5303.
23. Marteijn JA, Lans H, Vermeulen W, Hoeijmakers JHJ. Understanding nucleotide excision repair and its roles in cancer and ageing. Nat Rev Mol Cell Biol. 2014;15(7):465-481.

DNA NER - Transcription Coupled (TC-NER)

24. Hanawalt PC, Spivak G. Transcription-coupled DNA repair: two decades of progress and surprises. Nat Rev Mol Cell Biol. 2008;9(12):958-970.
25. Fousteri M, Mullenders LHF. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. Cell Res. 2008;18(1):73-84.

DNA Replication Stress

26. Ahuja AK, Jodkowska K, Teloni F, et al. A short G1 phase imposes constitutive replication stress and fork remodelling in pluripotent stem cells. Nat Commun. 2016;7:10660.
27. Desmarais JA, Hoffmann MJ, Bingham G, Gagou ME, Meuth M. Human embryonic stem cells and iPS cells display increased sensitivity to replication stress. Cell Cycle. 2012;11(3): 548-555.

Epigenetic Dysfunction

28. Nishino K, Toyoda M, Yamazaki-Inoue M, et al. DNA methylation dynamics in human induced pluripotent stem cells: epigenetic memory and reprogramming. Nat Cell Biol. 2011;13(5):541-549.
29. Nazor KL, Altun G, Lynch C, et al. Recurrent variations in DNA methylation in human pluripotent stem cells and their derivatives. Cell Stem Cell. 2012;10(5):620-634.
30. Lister R, Pelizzola M, Kida YS, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature. 2011;471(7336):68-73.

ERCC2 Function / Efficiency

31. Coin F, Marinoni JC, Rodolfo C, Fribourg S, Pedrini AM, Egly JM. Mutations in the XPD helicase gene result in XP and trichothiodystrophy. EMBO J. 1999;18(5):1357-1366.
32. Lehmann AR. TFIIH, DNA repair and transcription in human disorders. DNA Repair (Amst). 2001;1(2): 135-140.
33. Fassihi H, et al. Mutations in ERCC2 cause a spectrum of disorders. Nat Genet. 2016;48(2):198-203.

G1/S Transition

34. Sherr CJ, Roberts JM. CDK inhibitors: positive and negative regulators of G1-phase progression. Genes Dev. 1999;13(12):1501-1512.
35. Neganova I, Lako M. G1 to S phase cell cycle transition in embryonic stem cells. Cell Cycle. 2009;8(6): 715-723.

36. Singh AM, Dalton S. The cell cycle and cell fate in ESCs: G1 length as a critical determinant. Nat Rev Mol Cell Biol. 2009;10(11): 744-753.

G2/M Transition

37. Nurse P. Universal control mechanism regulating onset of M-phase. Nature. 1990;344(6266):503-508.
38. Lindqvist A, Rodríguez-Bravo V, Medema RH. The decision to enter mitosis: feedback and control. Trends Cell Biol. 2009;19(5):262-271.

Genome Instability / CIN

39. Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability-an evolving hallmark of cancer. Nat Rev Mol Cell Biol. 2010;11(3):220-228.
40. Bakhoum SF, Cantley LC. The multifaceted role of chromosomal instability in cancer and its microenvironment. Cancer Cell. 2018;34(6):954-965.
41. Sansregret L, Swanton C. The role of aneuploidy in cancer evolution. Nat Rev Cancer. 2017;17(1): 19-31.

γH2AX (gH2AX)

42. Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem. 1998;273(10):5858-5868.
43. Bonner WM, Redon CE, Dickey JS, et al. γH2AX and cancer. Nat Rev Cancer. 2008;8(12):957-967.
44. Mah LJ, El-Osta A, Karagiannis TC. γH2AX: a sensitive molecular marker of DNA damage and repair. Cancer Lett. 2010;327(1-2):123-133.

DNA Hypermethylation

45. Nazor KL, Altun G, Lynch C, et al. Recurrent variations in DNA methylation in human pluripotent stem cells and their derivatives. Cell Stem Cell. 2012;10(5):620-634.

DNA Hypomethylation / Demethylation

46. Lister R, Pelizzola M, Kida YS, et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. Nature. 2011;471(7336):68-73.
47. Nishino K, Toyoda M, Yamazaki-Inoue M, et al. DNA methylation dynamics in human induced pluripotent stem cells: epigenetic memory and reprogramming. Nat Cell Biol. 2011;13(5):541-549.

Marker Stability

48. Koyanagi-Aoi M, Ohnuki M, Takahashi K, et al. Differentiation-defective phenotypes in human induced pluripotent stem cells. Cell Stem Cell. 2013;12(6): 705-717.
49. Kim K, Doi A, Wen B, et al. Epigenetic memory in induced pluripotent stem cells. Nature. 2010;467(7313):285-290.
50. Mallon BS, Chenoweth JG, Johnson KR, et al. Stem cell lines and markers: practical characterization. Stem Cells. 2014;32(2): 461-473.

Mitochondrial Stress

51. Prigione A, Rohwer N, Hoffmann S, et al. Mitochondrial reprogramming and metabolic switch in iPSCs. Cell Stem Cell. 2011;9(4): 431-436.
52. Folmes CDL, Dzeja PP, Nelson TJ, Terzic A. Metabolic plasticity in stem cell homeostasis and differentiation. Cell Stem Cell. 2012;11(5): 596-606.

Nanog

53. Chambers I, Colby D, Robertson M, et al. Functional expression of Nanog. Cell. 2003;113(5):643-655.
54. Silva J, Smith A. Capturing pluripotency. Nature. 2008;453(7194): 419-421.

OCT3/4 (POU5F1)

55. Nichols J, Smith A. Naive and primed pluripotent states. Cell. 2009;138(4): 663-675.

SOX2

56. Avilion AA, Nicolis SK, Pevny LH, et al. Multipotent cell line requires Sox2 for maintenance. Genes Dev. 2003;17(1):126-140.
57. Masui S, Nakatake Y, Toyooka Y, et al. Pluripotency governed by Oct3/4 and Sox2. Nat Cell Biol. 2007;9(6):625-635.

ROS / Oxidative Stress

58. Armstrong L, Tilgner K, Saretzki G, et al. Human induced pluripotent stem cell lines show stress defense and mitochondrial aging signatures. Stem Cells. 2010;28(4): 741-752.
59. Saretzki G. Telomeres, mitochondria and oxidative stress in pluripotent stem cells. Philos Trans R Soc Lond B Biol Sci. 2014;369(1646):20130428.

The Hallmarks of Cancer

60. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
61. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;12(1):31-46.