DeepNEU©: Introducing aiCRISPRL, a hybrid AI stem cell and organoid simulation platform with broad gene editing capabilities and applications

Sally Esmail, and Wayne R Danter

123Genetix Inc., Windsor and London, Canada

ABSTRACT

CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated nuclease 9) provides powerful gene-editing tools that are applicable for gene therapy of a variety of diseases including, but not limited to cancer, rare diseases, and heart disease. In the current study, we first examined our artificial stem cell and organoid models that were generated by our literature validated DeepNEU platform from the perspective of gene-editing. We then evaluated the aiCRISPRL (aiCRISPR-Like) application of the DeepNEU platform by comparing the CRISPR-based gene-editing approach with the DeepNEU derived aiCRISPRL capabilities using artificial simulated HeLa cells (aiHeLa). We then evaluated the aiCRISPRL like capabilities of DeepNEU to introduce a series of specific mutations into the MutS homolog 2 (MSH2) gene to assess DNA Mismatch Repair (MMR). This approach permits a comparative assessment of CRISPR and aiCRISPRL technologies following the introduction of specific MSH2 mutations. When combined with our previous research the current data indicate that aiCRISPRL is an evolving AI platform technology that can be quickly and reliably deployed in gene therapy applications to complement wet-lab based gene-editing technologies.

INTRODUCTION

Overview of gene editing

Modern gene editing technologies like TALEN and CRISPR have revolutionized disease modeling, drug discovery, and gene therapy. The most widely known, used, and studied platform, CRISPR, uses a proprietary enzyme known as Cas9 to produce targeted double-stranded breaks in DNA [1]. These DNA breaks can then be used to insert, delete, or modify a gene or genes of interest. Multiple edits can also be introduced sequentially

to mimic a specific process like those that are thought to be typical of a progressive malignant disease like colorectal cancer (CRC) [1]. Following the introduction of the required edits, the cell's innate DNA repairs the modified site mainly by the non-homologous end joining (NHEJ) pathway with less contribution from the Homologous Repair (HR) pathway. CRISPR and other powerful gene-editing platforms continue to evolve in terms of accessibility, accuracy, and wider applicability [1-4].

Overview of the DeepNEU AI/ML platform

DeepNEU© is a literature validated AI/ML platform engineered to produce computer simulations of induced human organoids and pluripotent stem cells for studying rare and other diseases with a genetic basis. The primary purposes of DeepNEU simulations were to (1) better understand the disease pathophysiology that results from introducing specific gene mutations, (2) identify potential therapeutic targets and biomarkers, and (3) drive drug discovery and re-purposing. To date, this AI platform has been applied successfully to several mono and polygenetic diseases, viral pandemic preparedness and precision oncology. The DeepNEU platform also continues to evolve in terms of accessibility, accuracy, and wider applicability [5-13]. In the current project DeepNEU was used to create computer simulations of HeLa cell which we refer to as aiHeLa.

Comparison of CRISPR vs the DeepNEU simulated aiCRISPRL application

At a fundamental level, CRISPR and the DeepNEU platform are entirely different technologies that attempt to do many of the same things. The modern CRISPR era began in 2009 [14] and is a widely validated wet lab technology, while DeepNEU is a more recent, [5] advanced and literature validated artificial intelligence computational platform. Both approaches have their own limitations. For example, in CRISPR, the enzyme used to cut DNA is not perfect and the cellular DNA repair process is not completely error-free. These CRISPR limitations can result in unwanted off-target effects, scars and these errors can be inherited. In addition, current wet lab gene-editing technologies are expensive and time-consuming with varying success rates. The current version of DeepNEU(v7.0) also has a few limitations. First, the current database from which the AI learns is incomplete but growing. The current database contains gene-gene and gene-protein relationships that account for just over 25% of the human genome. Although ~75% of the human genome remains to be included, the database is updated regularly with the latest information. The goal is

to reach at least 99% representation over the next two to three years. Of note, the simulations reached a stable ability to learn and generalize well since reaching representation of 20% of the genome. Secondly, the system predictions have been validated against a hold-out sample of previously unseen peer-reviewed wet lab results. By combining unsupervised learning with early stopping and big data for regularization, this platform is increasingly moving towards wet lab testing as employed for CRISPR and other gene-editing platforms.

Based on the available data we believe that aiCRISPRL (aiCRISPR-Like) and gene-editing technologies like CRISPR have important similarities. For example, both can create and evaluate knock-in/Gain of Function (GOF) and knock-out/Loss of Function (LOF) mutations. They both can also evaluate a broad range of single and multiple gene modifications either together or sequentially. Importantly, the proposed computational approach avoids unwanted off-target effects and is both time and cost-effective but lacks the robust validation of the wet lab CRISPR technology.

In the current study, we first examine our previous artificial stem cell and organoid research from the perspective of gene-editing [5-13]. We then extended our previous work by evaluating the aiCRISPRL application of the DeepNEU platform by comparing a CRISPR-based gene editing approach with the aiCRISPRL approach applied to artificial HeLa cells (aiHeLa) exposed to different MSH2 mutations to assess DNA Mismatch Repair (MMR). This approach should permit direct comparison of the two gene-editing approaches following introduction of specific mutations in the MSH2 gene.

METHODS

The DeepNEU platform is an evolving literature-validated unsupervised machine deep machine learning platform [5-13]. Here we are presenting the upgraded database of DeepNEU (v7.0). this version contains the previous database (v6.6), in addition to an upgrade in the form of 67 new genotypic and phenotypic concepts and ~2000 new relationships.

The initial goal of this project was to generate simulated HeLa cells (aiHeLa) by applying aiCRISPRL gene editing of the simulated version of induced pluripotent stem cells (aiPSC) genome to introduce consensus gene mutations typical of HeLa cells [15-18]. Briefly, the

values of the high dimensional input or initial state vector (N=5501 concepts) of the aiPSC model, were all set to zeros except for the four transcription factors OCT3/4, SOX2, KLF4 and cMYC. These four factors were given a value of +1 indicating that they were turned on for the first iteration and then allowed to evolve over successive iterations until a new system wide steady state is achieved. The detailed methodology for generating aiPSC simulations has been reported in detail previously [5-13].

Previous versions of the DeepNEU platform have successfully simulated a broad range of aiPSC derived cell types including neural stem cells, cardiac myocytes, skeletal muscle cells, lung cells, brain cells, ovarian cancer cells, lung adenocarcinoma cells, and natural killer cells (NK) [5-13]. Using a similar approach and a consensus gene mutational profile[15-18], we used the gene-editing capabilities of aiCRISPRL to create simulations (aiHeLa) that are most like original HeLa cells. Importantly, except for the simulated HPV-18 infection, the aiHeLa cells were created to be devoid of all external contaminations that plague HeLa variant clones that have emerged since 1951.

Simulation of HeLa cells (aiHeLa)

Once the aiPSCs were validated against the current literature they were transformed by aiCRISPRL into aiHeLa cells using a gene mutational profile derived from the peer-reviewed literature [15-18]. This aiHeLa mutation profile included GOF mutations in CD24, CD44, CD95(FasR), HPV-E2, 5, 6, and 7. Aneuploidy, an important feature of HeLa cells, was also locked ON and the cells were maintained in simulated DMEM/F12 media supplemented with doxycycline. This information is summarized in Table 1 below.

Table 1: Summary of aiPSC and aiHeLa cell simulation protocol

Concept	Recipe	Component/Edits
aiPSC	Yamanaka transcription factors	OCT4, cMYC, KLF4 and SOX2 – GOF gene edits
aiHeLa	Mutational profile	CD24, CD44, CD95(FasR), HPV-E2,5,6,7– GOF gene edits, Aneuploidy turned ON

DMEM/F12 media	Basic media	[eNa+], [ePyruvate], [eZinc], Choline, FGF2/bFGF, Folic Acid, HEPES, HSA/Albumin, Hypoxanthine, Inositol, KSR, Nicotinamide, ROCK1/2 inhibitor, VitB12/Cobalamin, VitB6/Pyridoxine
DMEM/F12 media	Supplements	Doxycycline turned ON
Environment	Settings	[O2]=21%, [CO2]=5%, Temperature=37 º C - locked ON
Patient factors	Age, Smoking	31 years, smoker – Locked ON

Outcome predictions from the HeLa simulations (aiHeLa) were directly compared with published data [15-18]. Furthermore, expression values ≥ 0 were classified as expressed or upregulated for genes/proteins or present for phenotypic features while values < 0 were classified as downregulated, not expressed, or absent.

To evaluate the ability of the DeepNEU platform to simulate HeLa cells, a consensus feature profile was created from the published literature [15-18]. Each feature was identified in at least two published references. The final HeLa cell feature profile (N=9) includes ALDH1, Aneuploidy, CD24, CD44, CD95(FasR), HPV-E6, HPV-E7, Sox2 and Telomerase. The accuracy of simulation predictions was compared to the published literature using the unbiased binomial test. A test p-value <0.05 was used to reject the null hypothesis (H^0) that the simulated HeLa cell (aiHeLa) profile could not reproduce the documented wet-lab results.

aiCRISPRL- simulated gene editing of the MSH2 gene in aiHeLa cells

Following the successful creation and literature validation of the aiHeLa simulations, we turned our attention to modifying the MSH2 gene to produce a range LOF mutations. To allow a direct comparison with published data we followed the approach and analysis presented in [19]. These authors compared wildtype MSH2 to seven specific MSH2 LOF

mutations in HeLa cells that were created by CRISPR gene editing. Our approach is modified from [19] and summarized in Table 2.

Table2: Summary of aiCRISPRL gene editing of MSH2 conducted in simulated HeLa cells (aiHeLa)

Simulation Name	aiCRISPRL Gene Editing	MSH2-% (Scaled*)
aiHeLa	None (WT)	100 (+1)
aiHeLa-MSH2-KO	MSH2-deleted	0 (-1)
aiHeLa-RD03	MSH2 G674R1	11 (-0.780)
aiHeLa-RH07	MSH2 G674R2	5.2 (-0.896)
aiHeLa-DA02	MSH2 G674D1	4.8 (-0.904)
aiHeLa-DC08	MSH2 G674D2	3.4 (-0.932)
aiHeLa-AG11	MSH2 G674A1	30 (-0.400)
aiHeLa-AH07	MSH2 G674A2	35 (-0.300)

Modified from [19], 2019, (Scaled*) - the % MSH protein expression was converted to the -1 to +1 range used as inputs to the DeepNEU/aiCRISPRL platform

Specific MSH2 LOF mutations were assigned a simulation name and programmed to reproduce the associated changes in the amount of MSH2 protein expression as per Table 2.

Loss of Function (LOF) Scores

In their paper [19] defined and validated a LOF score for a group of seven CRISPR edited MSH2 mutations plus WT that they evaluated. To develop this score, the HeLa cells were treated with 6-Thioguanine (6-TG). Their MOA-based reasoning led them to conclude that the cells with intact MMR pathways should be sensitive to 6-TG exposure while the cells with defective pathways (dMMR) should show varying degrees of resistance as measured by cell death. Their LOF score was calculated from the Log (base2) of the ratio of 6-TG to placebo treatment effects in each of the gene-edited and WT cells. Positive LOF scores imply deleterious mutations while negative scores suggest more neutral mutations. Analysis of their data revealed a Pearson correlation r of 0.770 between the severity of the MSH2 mutations and HeLa cell death.

Our aiCRISPRL generated LOF scores were modified from [19] in that the amount of HeLa cell apoptosis was substituted for MSH2 mutation severity since 6-TG primarily induces cell death through apoptosis [20]. In addition, the DeepNEU input scaling range from -1 to +1 results in some negative values for aiHeLa cell apoptosis necessitating a reversal of numerator and denominator to derive the ratio used to produce the LOF score. As in [19], positive LOF scores imply deleterious mutations, and negative scores indicate typically more neutral mutations. In the present project the correlation between the LOF score and MSH2 protein expression was evaluated by calculating the Pearson correlation coefficient (r).

RESULTS

The updated version of DeepNEU (v7.0) contains 5501 concepts and 50437 nonzero relationships compared to the previous DeepNEU(v6.6) database that included 5434 concepts and 48,487 nonzero relationships. While this represents an increase of just 67 concepts, there are almost 2000 new relationships primarily related to HeLa cells and mismatch DNA repair (MMR) pathways. The current database (v7.0) represents ~25% of the human genome and is growing daily. In addition, a detailed review of the relationships revealed that the pretest probability of a positive outcome is 0.660 and the probability of a negative prediction is therefore 0.340. These values were used to optimize the binomial test by eliminating system biases prior to its use.

On review of our previous publications [5-13], the DeepNEU platform successfully used gene editing to create single-gene mutations, multiple mutations and even edit the entire genome of the SARS-CoV-2 and NIPAH viruses. These edits created both loss of function (LOF) and gain of function (GOF) mutations. A summary of these data from previously published DeepNEU research projects from the perspective of successful gene editing is presented in Table 3 below.

Table 3: aiCRISPRL-Like gene-editing applications of the DeepNEU platform

Date	aiStem Cell	aiNSC	aiSkMC	aiLUNG	aiOrganoid	Disorder	LOF Mutations	GOF Mutations	Reference
2019	Yes	Yes	No	No	No	RETT Syndrome	MeCP2	None	1
2019	Yes	No	Yes	No	No	IOPD	GAA	None	2
2020	Yes	No	No	No	No	SARS-CoV-2 #1	Whole genome	Whole genome	3
2021	Yes	No	No	No	Whole Brain	Alzheimer's Disease	None	APO-E	4
2021	Yes	No	No	No	Lung	SARS-CoV-2 #2	Whole genome	Whole genome	5
2021	Yes	No	No	No	Lung	LUAD	TP53	EML4-ALK fusion, Amplicon 11q13, NFKBIA, NKX2-1	6
2021	Yes	No	No	Yes	No	SARS-CoV-2 #3	Whole genome	Whole genome	7
2021	Yes	No	No	Yes	Whole Brain	MLD	ARSA	None	8
2021	Yes	No	No	No	No	HGSOC	Many	Many	9
2022	Yes	No	No	No	Whole Brain	NIPAH encephalitis	Whole genome	Whole genome	Pending

aiCRISPRL-derived aiHeLa Cells Simulations

The aiCRISPRL gene modifications to the aiPSC described above resulted in aiHeLa cells with a profile consistent with the literature-derived nine feature HeLa profile including ALDH1, Aneuploidy, CD24, CD44, CD95, HPV-E6,7, Sox2 and Telomerase. The probability that all these markers were expressed by chance alone is 0.024 based on the unbiased binomial test. The H^0 can therefore be rejected in favor of the alternate hypothesis (H^1) that the simulated HeLa cell (aiHeLa) profile can accurately reproduce published wet lab results. These results are summarized below in Figure 1.

LOF Score Analysis

Analysis of the relationship between MSH2(scaled) and the calculated LOF score revealed a strong, negative correlation with a Pearson r value of -0.900. Based on the sample size of 8 and a two-tailed critical r value of 0.765, the probability that this correlation occurred by chance alone is <0.01. Allowing for the scaling modifications to

our MSH2 LOF score, the strong negative correlation is consistent with [19] in that more positive (or less negative in the case of aiCRISPRL) LOF scores imply deleterious mutations while more negative scores suggest neutral mutations. These results are presented in Table 4 and Figure 2.

Table 4 Pearson correlation (r) between MSH2 protein and LOF Score = -0.900

Model-Average, N = 3	**MSH2***	**6-TG****	**Placebo*****	**Ratio#**	**LOF Score##**
aiHeLa_Cell (24/11/21)	+1	-0.746	-0.615	0.825	-0.278
aiHeLa_Cell (24/11/21)-AG11	-0.400	-0.643	-0.603	0.938	-0.093
aiHeLa_Cell (24/11/21)-AH07	-0.300	-0.641	-0.623	0.973	-0.040
aiHeLa_Cell (24/11/21)-DA02	-0.904	-0.344	-0.389	1.129	0.175
aiHeLa_Cell (24/11/21)-DC08	-0.932	-0.358	-0.368	1.028	0.040
aiHeLa_Cell (24/11/21)-KO	-1	-0.366	-0.426	1.164	0.220
aiHeLa_Cell (24/11/21)-RD03	-0.780	-0.422	-0.450	1.066	0.092
aiHeLa_Cell (24/11/21)-RH07	-0.896	-0.341	-0.402	1.180	0.238

Table 4 Legend: MSH2*= MSH2 gene expression scaled between -1 to +1 for input to the DeepNEU platform; 6-TG** = the predicted effect of 6-TG on aiHeLa apoptosis; Placebo*** = the predicted effect of Placebo on aiHeLa apoptosis; Ratio$^{\#}$= ratio of Placebo effect over 6-TG effect; LOF Score$^{\#\#}$= Log_2(Ratio) as per [19]

DISCUSSION

Our previous successful gene editing applications of the DeepNEU platform are summarized in Table 1 above [5-13]. In the current project, we have extended those results by simulating HeLa cells and then introducing variable LOF mutations into a specific DNA damage repair gene (MSH2) in the aiHeLa cells. Based on a literature derived 9 element feature profile [15-18], our analysis of the data indicates that the aiCRISPRL application has successfully simulated a version of the original HeLa cell line (aiHeLa). Since 1951, the original HeLa cell genome has diverged to form numerous unique clones across the globe. These widespread changes in the HeLa genome have been the result of sequential infections from multiple bacteria, fungi, and viruses aided by more than 70 years of mutational pressures and rapid cell division. HeLa cells in culture have a doubling time of just 23 hours. As a result, available HeLa cells have little in common with the original cell line [21-24].

An important advantage of our DeepNEU/aiCRISPRL application for engineering aiHeLa cells allows us to avoid any unwanted external contamination from chemicals and infective organisms using a standardized culture medium. The only source of infection in these simulations comes from the simulated HPV-18 that generates E proteins 2, 5, 6, and 7. The impact of our early stopping regularization on system evolution is to reduce the impact of protracted time on frequent DNA replication and rapid cell division. The use of a consensus approach to mutational profile was also designed to eliminate the development of any mutations resulting from the combination of time, laboratory conditions, and contamination by multiple infecting organisms. Taken together these factors suggest that the simulations have produced a generic HPV-18 induced cervical cancer in a 31-year-old woman smoker and as a result, these simulations are likely to be more like the original HeLa cell type than current divergent clones.

Once the aiHeLa cell simulations were validated using the previously unseen published literature, the DeepNEU/aiCRISPRL platform was used to modify the aiHeLa cell MSH2 gene. This was conducted by introducing several (N=7) LOF mutations to produce a

range of negative effects on MSH2 protein expression and then comparing these LOF mutations with the wildtype aiHeLa. The relationship between LOF mutations and the resulting MSH2 protein levels was scaled from the format in Table 3, [19] to the

DeepNEU input range where -1 = deletion of the MSH2 gene resulting in complete loss of protein and +1 = wild type gene status and normal levels of MSH2 protein.

The impact of these LOF mutations on aiHeLa cell DNA damage repair was evaluated by assessing the impact of 6-Thioguanine (6-TG) treatment on both WT and mutated aiHeLa cells as described in [19]. 6-TG is a thiopurine prodrug metabolized to its active form in the liver. The active form undergoes further metabolism to produce thioguanine nucleotides (6-TGNs) that can be incorporated into RNA and DNA synthesis as false purines resulting in potentially lethal DNA mutations. While the MMR process recognizes these mutations, it is unable to repair them resulting in replication arrest and apoptosis. Additional cytotoxicity from 6-TGNs is the result of inhibiting the RAC1 protein that regulates the diverse downstream signals of the RAC1-VAV pathway in various cancer cells [26].

Consistent with the data from [19] we were able to confirm the effect of MSH2 protein loss on HeLa cell mortality as measured by the degree of apoptosis. When we analyzed the relationship between MSH2(scaled) and the calculated LOF score a strong, negative correlation with a Pearson r value of -0.900 was revealed (N=8, critical r = 0.765, $p<0.01$). This finding confirms that the effectiveness of 6-TG to induce apoptosis in aiHeLa cells is dependent on a functioning MSH2 dependent MMR machinery. Importantly, aiCRISPRL editing of the MSH2 gene can accurately reproduce specific mutations and the LoF scores reported in [19]. In addition, this important relationship appears to be stronger for the aiCRISPRL editing case (i.e., Pearson r =-0.900 vs 0.770).

Conclusions and Future Considerations

In this study, we successfully employed the DeepNEU platform to simulate aiHeLa cells that accurately resembled the original and uncorrupted immortalized HeLa cell line. We then evaluated the DeepNEU derived aiCRISPRL like capabilities to introduce a series of specific LOF mutations into the MSH2 gene. We chose to study the MSH2 gene because it is a critical component of the MMR DNA repair pathway and it could be evaluated by treating the affected HeLa cells with the cytotoxic prodrug, 6 Thioguanine (6-TG), and observing the degree of resulting apoptosis. The severity of the loss of function mutations in MSH2 was estimated from the LOF score. This score was calculated in a manner like

that reported in [19] and produced data confirming a highly significant inverse correlation between MSH2 protein levels and aiHeLa cell apoptosis. This methodology has permitted the direct comparison of CRISPR-Cas9 and aiCRISPRL technologies for introducing specific MSH2 mutations into the HeLa/aiHeLa genome and while the technologies are different, they are directly comparable. Furthermore, like CRISPR-Cas9, aiCRISPRL can be used to create single, multiple as well as a sequential (LOF and GOF mutations (see Table 1).

CRISPR-Cas9 is a mature and impressive in vitro gene-editing technology with an improving success rate while aiCRISPRL is a specific application of an evolving AI platform technology that can be quickly and reliably deployed. Like other AI simulation technologies, DeepNEU requires substantial amounts of validated data from multiple sources. The current DeepNEU network is composed of >3X10^7 artificial neurons and the database contains relationship data for >25% of the human genome. Our objective is to obtain relationship data for ~99% of the human genome over the next 2-3 years.

To date, the DeepNEU platform has relied on previously unseen and published wet-lab data to function as a hold-out sample for the validation of system predictions. We are now embarking on a program of additional wet-lab experiments to further validate the DeepNEU platform, beginning in the coming months.

Finally, we have evidence that the DeepNEU platform has recently entered the domain of Wise Learning (WL) as it relates to health care. The WL process defined by Groumpos in 2016, represents the next evolutionary step in AI that combines Fuzzy Cognitive Map simulations, with data from multiple experts and a Generic Decision-Making System (DMS). The WL process should also explore available learning algorithms including Deep Learning (DL) methods when available [25]. While the DeepNEU platform continues to evolve, as of this writing, the current version (7.0) meets all these Wise Learning stated criteria.

FIGURE LEGENDS

Figure 1: aiCRISPRL HeLa Cell simulations profile. aiHeLa Cell Marker expression. The vertical y-axes represent the semiquantitative levels of markers that are estimated by DeepNEU relative to an arbitrary baseline where 0 = baseline, +1 = maximum expression or presence and -1 = minimal expression level or presence. The horizontal x-axis represent the aiHeLa markers being simulated. Data represent means of three experiments ± 95% confidence interval. All p values from the Welch's t test. *p<0.01

Figure 2: aiCRISPRL analysis of MSH2 Mutations in aiHeLa Cells. Analysis of the relationship between MSH2(scaled) and the calculated LOF score. Analysis of the relationship between MSH2(scaled) and the calculated LOF score revealed a strong, negative correlation with a Pearson r value of -0.900, based on the sample size of 8 and a two-tailed critical r value of 0.765. Data represent means of three experiments ± 95% confidence interval. All p values from the Welch's t test. *p<0.01

FIGURE 1

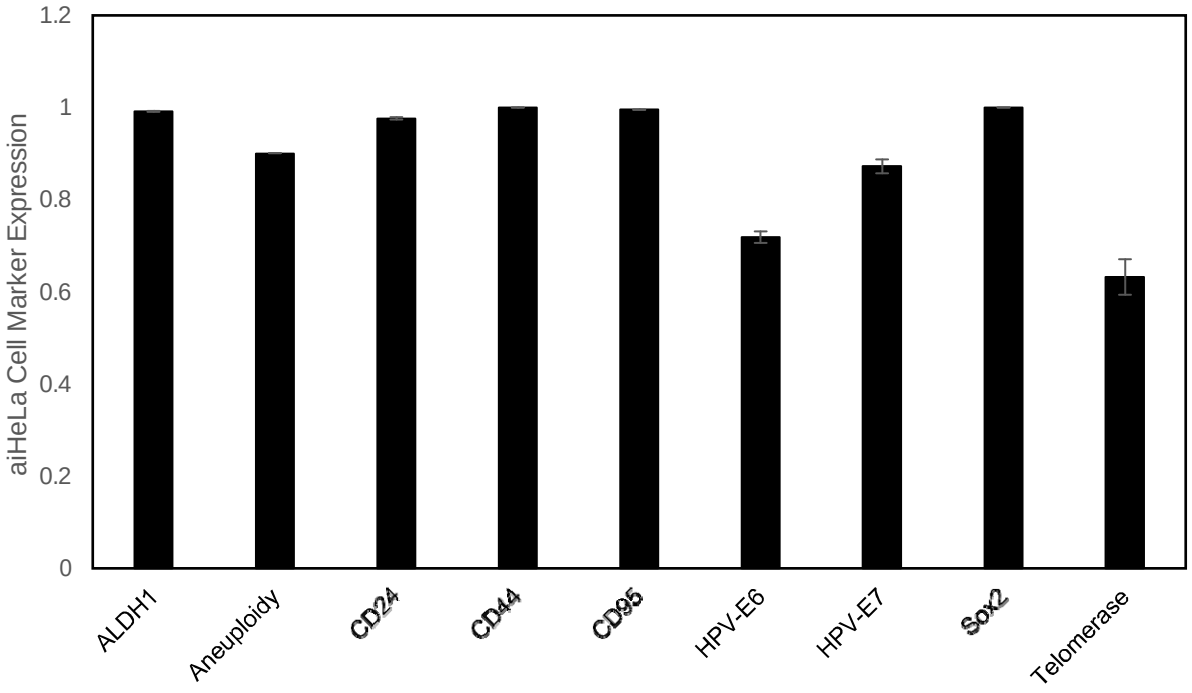
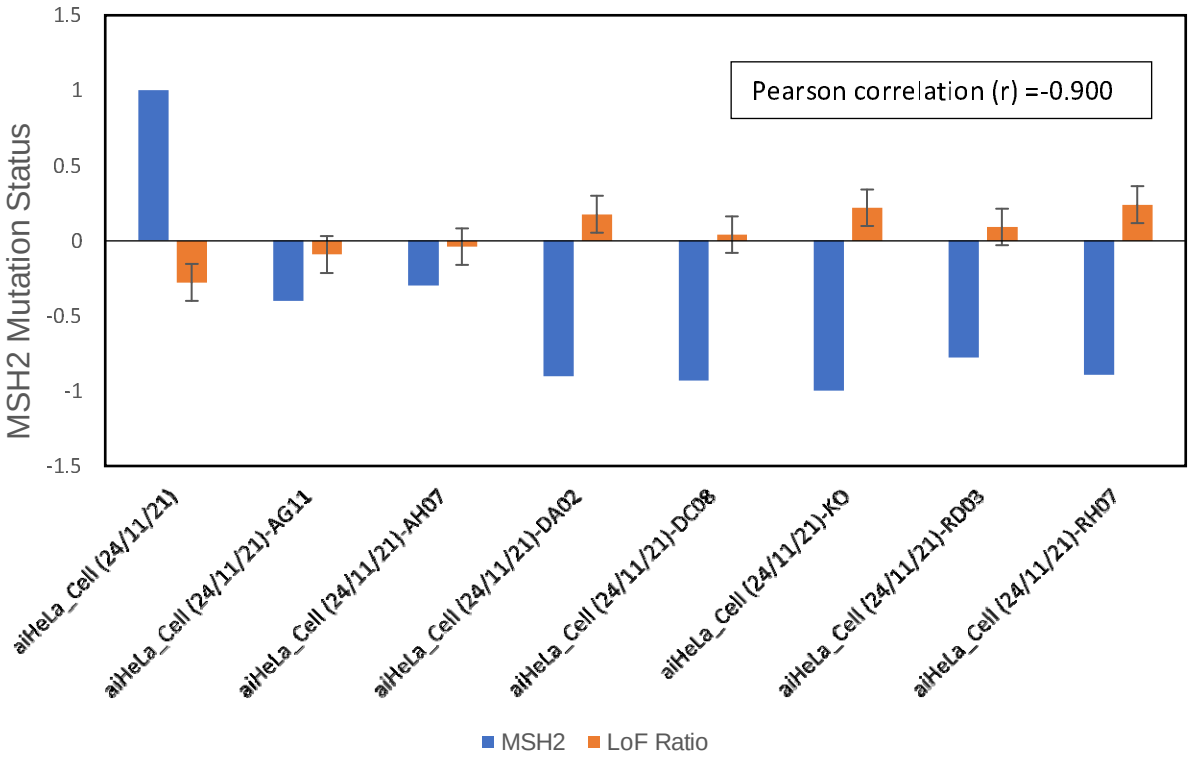


FIGURE 2



REFERENCES

1 Liu, M., *et al.* (2021) Global detection of DNA repair outcomes induced by CRISPR–Cas9. *Nucleic acids research* 49, 8732-8742

2 Gopal, S., *et al.* (2020) Exploiting CRISPR Cas9 in three-dimensional stem cell cultures to model disease. *Frontiers in Bioengineering and Biotechnology* 8, 692

3 Li, H., *et al.* (2020) Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances, and prospects. *Signal transduction and targeted therapy* 5, 1-23

4 Vlachogiannis, G., *et al.* (2018) Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* 359, 920-926

5 Danter, W.R. (2019) DeepNEU: cellular reprogramming comes of age–a machine learning platform with application to rare diseases research. *Orphanet journal of rare diseases* 14, 1-13

6 Esmail, S. and Danter, W.R. (2022) Stem-cell based, machine learning approach for optimizing natural killer cell-based personalized immunotherapy for high-grade ovarian cancer. *The FEBS Journal* 289, 985- 998

7 Esmail, S. and Danter, W.R. (2019) DeepNEU: artificially induced stem cell (aiPSC) and differentiated skeletal muscle cell (aiSkMC) simulations of infantile onset POMPE disease (IOPD) for potential biomarker identification and drug discovery. *Frontiers in cell and developmental biology*, 325

8 Esmail, S. and Danter, W. (2021) Viral pandemic preparedness: A pluripotent stem cell-based machine- learning platform for simulating SARS-CoV-2 infection to enable drug discovery and repurposing. *Stem cells translational medicine* 10, 239-250

9 Esmail, S. and Danter, W.R. (2021) NEUBOrg: Artificially Induced Pluripotent Stem Cell-Derived Brain Organoid to Model and Study Genetics of Alzheimer's Disease Progression. *Frontiers in aging neuroscience* 13, 53

10 Esmail, S. and Danter, W.R. (2021) Lung organoid simulations for modelling and predicting the effect of mutations on SARS-CoV-2 infectivity. *Computational and Structural Biotechnology Journal* 19, 1701- 1712

11 Esmail, S. and Danter, W.R. (2021) A novel artificial lung organoid for simulating a patient derived adenocarcinoma of lung for personalized oncology. *medRxiv*

12 Esmail, S. and Danter, W. (2021) DeepNEU: a machine learning platform for simulating Cytokine Storm and Coagulopathy that complicate severe COVID-19 to enable targeted drug repurposing.

13 Esmail, S. and Danter, W.R. (2021) Artificially Induced Pluripotent Stem Cell-Derived Whole-Brain Organoid for Modelling the Pathophysiology of Metachromatic Leukodystrophy and Drug Repurposing. *Biomedicines* 9, 440

14 Wiedenheft, B.*, et al.* (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* 17, 904-912

15 Landry, J.J.*, et al.* (2013) The genomic and transcriptomic landscape of a HeLa cell line. *G3: Genes, Genomes, Genetics* 3, 1213-1224

16 Mittelman, D. and Wilson, J.H. (2013) The fractured genome of HeLa cells. *Genome biology* 14, 1-4 17 Wang, L.*, et al.* (2014) Enrichment and characterization of cancer stem□like cells from a cervical cancer cell line. *Molecular medicine reports* 9, 2117-2123

18 Gutiérrez-Hoya, A.*, et al.* (2019) Cervical cancer cells express markers associated with immunosurveillance. *Journal of immunology research* 2019

19 Hayashida, G.*, et al.* (2019) Differential genomic destabilization in human cells with pathogenic MSH2 mutations introduced by genome editing. *Experimental Cell Research* 377, 24-35

20 Li, H.*, et al.* (2020) Transcriptomics analysis of the tumor-inhibitory pathways of 6-Thioguanine in MCF-7 cells via silencing DNMT1 activity. *OncoTargets and therapy* 13, 1211

21 Masters, J.R. (2002) HeLa cells 50 years on: the good, the bad and the ugly. *Nature Reviews Cancer* 2, 315-319

22 Tang, L. (2019) Investigating heterogeneity in HeLa cells. *Nature Methods* 16, 281-281

23 Liu, Y.*, et al.* (2019) Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nature biotechnology* 37, 314-322

24 Kniss, D.A. and Summerfield, T.L. (2014) Discovery of HeLa cell contamination in HES cells: call for cell line authentication in reproductive biology research. *Reproductive sciences* 21, 1015-1019

25 Groumpos, P.P. (2016) Deep learning vs. wise learning: a critical and challenging overview. *IFAC- PapersOnLine* 49, 180-189

26 Zakerska-Banaszak, O.; Lykowska-Szuber, L.; Walczak, M.; Zuraszek, J.; Zielinska, A.; Skrzypczak-Zielinska, M. Cytotoxicity of Thiopurine Drugs in Patients with Inflammatory Bowel Disease. Toxics 2022, 10, 151