



Figure 1. Distribution of ATMP Clinical Trials by Therapy Type (n=130).Figure 2. Global ATMP Trials by Developmental Phase

2380 PANDORA: A multiparametric framework for sensitive detection of insertional genotoxicity

Riccardo Pizzichemi^{1,2}, Francesco Gazzo^{1,2}, Alice Bonomi¹, Barbara Gega¹, Sergio Arevalo Diaz¹, Fabrizio Benedicenti¹, Marco Masseroli², Eugenio Montini¹

¹San Raffaele Telethon Institute for Gene Therapy (SR-Tiget) IRCCS San Raffaele Scientific Institute, Milan, Italy,

²Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy

Introduction

In hematopoietic stem cell gene therapy, integrating vectors insert therapeutic transgenes into semi-random genomic positions to correct genetic defects. However, vector integration can potentially result in proto-oncogene activation or tumor suppressor inactivation and trigger cancer. Current approaches used to evaluate insertional mutagenesis-mediated genotoxicity rely mainly on two readouts: Common Insertion Sites, indicating genetic selection and the enrichment of Insertion Sites (ISs) targeting specific genes, and clonal dominance, defined as the progressive expansion of individual clones exceeding 30% of clonal abundance.

Methods

To increase the predictive power of genotoxicity detection, we designed a new analytical method, PANDORA (PAN Detection Of insertional mutagenesis), which overcomes the limited sensitivity of CIS and clonal dominance analyses by incorporating three additional complementary readouts:

1. Abnormal exon targeting frequency: assessing enrichment of ISs in specific gene exons by comparing observed versus expected exonic targeting frequency based on the exon-to-gene length ratio using a chi-square test.
2. Rank-based clonal abundance evaluation: Increased clonal abundance of ISs targeting specific genes quantified by comparing their percentile rank distribution to the average percentile rank across all genes using a Mann–Whitney U test.
3. IS orientation bias: Bias in IS orientation relative to gene transcription, suggestive of promoter- or enhancer-mediated activation, evaluated by testing deviations from the expected 50:50 distribution using a chi-square test.

Multiple testing correction was applied using the Benjamini Hochberg FDR correction.

Results

In our initial analysis we applied PANDORA on >6.5 million ISs from blood cells of 50 patients enrolled in three different gene therapy trials: a g-retroviral vector (g-RV) clinical trial for Wiskott-Aldrich syndrome (WAS) (370,000 ISs, n=10 Pts), in which insertional oncogenesis was reported in most patients; a lentiviral vector (LV) clinical trial for X-linked severe combined immunodeficiency (SCID-X1; 3.9 million ISs, n = 8), in which several patients developed clonal dominance associated with premature transcriptional termination and aberrant splicing, caused by activation of a cryptic splice site within the chicken hypersensitive site-2 (cHS4) insulator sequence embedded in the vector LTR, leading to loss of micro RNA-mediated regulation and pathological gene overexpression (e.g. HMGA2); lastly, a LV clinical trial for metachromatic leukodystrophy (MLD, 2.3 M ISs, n=32), containing a moderate PGK promoter in internal position and where no insertional mutagenesis events have been reported to date.

PANDORA identified, across the 3 gene therapy trials, 2,470 putative culprits of insertional mutagenesis: 22% (544) by exonic enrichment, 77.4% (1,912) by abnormal clonal abundance, and 0.56% (14) by IS orientation bias. Interestingly, exonic insertions were biased toward the last exon and 3' untranslated region (3' UTR), enriched in microRNA target sites whose disruption leads to microRNA control impairment and gene overexpression.

Restricting the analysis to genes recurrently identified in two or more patients we found 51 culprits in γ -retroviral WAS, 294 in LV-cHS4 SCID-X1, and 113 in LV-PGK MLD, indicating that vector-mediated fitness modulation is widespread and involves several genes. Compared with conventional analyses, PANDORA detects orders of magnitude more candidate genes.

Recurrent targets included HMGA2, MECOM, TAL1 in γ -RV WAS; HMGA2, IKZF1, KMT2D, ETS1, NF1, RB1, and BRCC3 in LV-cHS4 SCID-X1; and KDM2A, IKZF1, PTEN, KLF5, NF1 and BRCC3 in LV-PGK MLD. The functional composition of these genes differs by vector design, consistent with distinct genotoxic mechanisms: enhancer activation in γ -RV WAS, complex splicing and truncation mediated effects in LV-cHS4 SCID-X1, and a more constrained but detectable fitness impact in LV-PGK MLD. Genes identified by PANDORA are strongly enriched for cell-cycle and cancer-related functions compared with the genomic background, showing a significantly higher proportion ($p < 0.001$, by chi-square test) among PANDORA-identified candidates than among all integration-targeted genes across gene therapy trials.

The number of PANDORA candidate genes per million IS differed significantly across trials, with LV-cHS4 SCID-X1 showing the highest burden (~150–300 genes/M IS), followed by γ -RV WAS (~50–200 genes/M IS), and markedly lower levels in LV-PGK MLD (~10–30 genes/M IS), highlighting vector-dependent genotoxic risk.

Conclusion

PANDORA provides a scalable and sensitive multiparametric framework to quantitatively assess and stratify insertional genotoxicity across vector designs, enabling improved safety evaluation, informed vector optimization, and deeper mechanistic insight into vector–host genome interactions in gene therapy.

2381 Mechanistic and variant-based patient stratification in CSF1R-related disorder to inform clinical trial design: a framework for rare inherited diseases

Yedda Li¹, Florian Eichler¹

¹*Massachusetts General Hospital, Neurology, Boston, United States*

Introduction

Monoallelic *CSF1R*-related disorder (CRD) is a fatal, primarily adult-onset leukodystrophy caused by pathogenic variants in the colony stimulating factor 1 tyrosine kinase receptor (CSF1R). Currently, allogeneic hematopoietic stem cell transplantation (HSCT) is the only disease-modifying therapy available, and *ex vivo* hematopoietic stem cell (HSC)-directed precision genome editing is emerging as a promising alternative therapy. However, current clinical decision-making and future clinical translation are both impeded by an incomplete understanding of how the wide phenotypic spectrum associated with *CSF1R* variants can be stratified in a clinically meaningful way. This limits patient selection, prioritization of gene editing strategies, and definition of appropriate outcome measures for clinical trials. More broadly, this problem reflects a fundamental challenge in rare genetic diseases, where insufficiently integrated genotype-phenotype data hampers rational trial design despite rapid advances in gene and cell therapies.

Methods

Using a multi-source curation approach incorporating published reports and clinical databases, we established an international database of *CSF1R* variants spanning symptomatic CRD, asymptomatic CRD, and non-CRD neurodegenerative phenotypes, including Alzheimer disease, frontotemporal dementia, Parkinson disease, multiple sclerosis, and amyotrophic lateral sclerosis, among others. Variants were annotated by protein domain localization, predicted nonsense-mediated decay (NMD) susceptibility, disease penetrance, and age of onset, then analyzed to define clinically relevant, mechanism-based genotype-phenotype relationships.