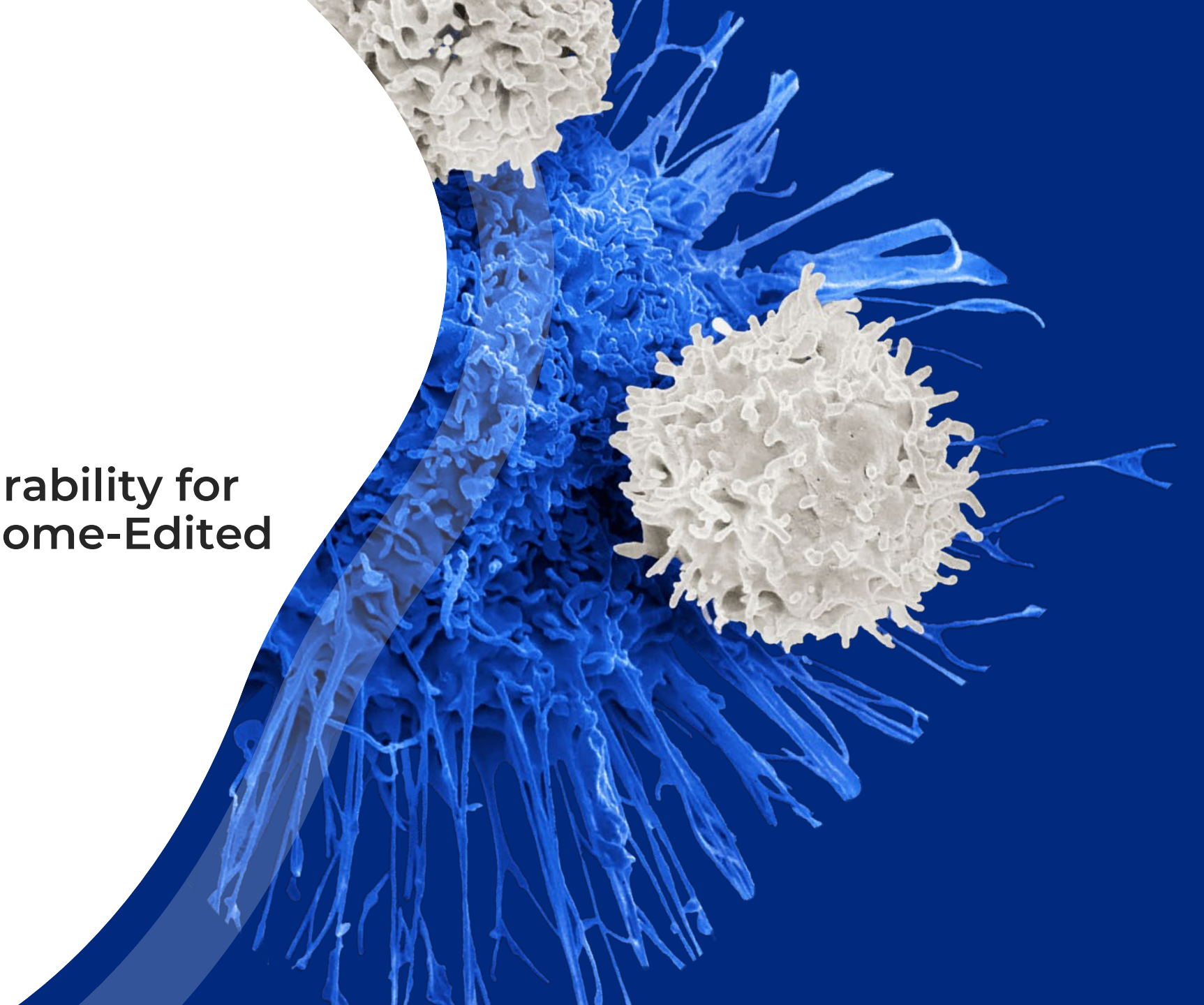
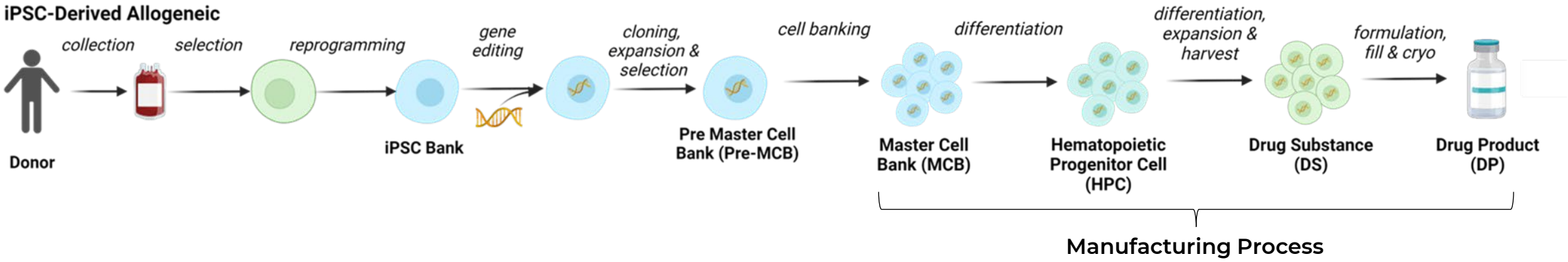
CENTURY THERAPEUTICS

Case Study in Comparability for an iPSC-Derived, Genome-Edited Cell Therapy Product

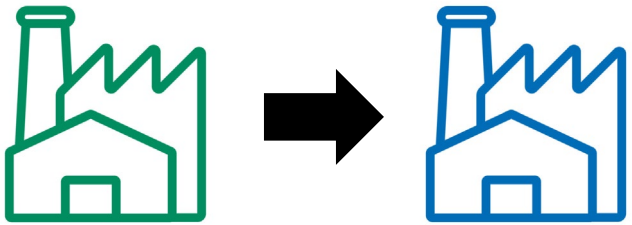
Jennifer L. Dashnau, PhD, MBA
IABS
14 Nov 2024



New facility introduction during Phase 1 is considered a change requiring comparability evaluation



Manufacturing Change



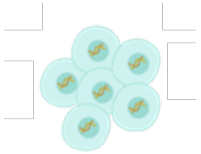
Facility fit and risk assessment performed to describe change and determine risk:

- New facility, same single-use process
- Scope included HPC to DP (MCB to HPC not included)
- Limited changes:
  - Minor process choreography differences
  - Instrument model differences (new vs discontinued model)
  - Bioreactor design update (generation 2 vs 1)
  - Raw material changes (vendor changes)

Risk to product quality, safety, and efficacy considered low

Prospective comparability study structured as comparison of new facility batches to historical batches

Master Cell Bank (MCB)



Hematopoietic Progenitor Cell (HPC)

Drug Substance (DS)

Drug Product (DP)

MCB 1	*HPC Demo 1*	*DS Demo 1*	*DP Demo 1*
		DS Demo 2	*DP Demo 2*
		DS Shakedown	DP 1 Shakedown
	HPC Engineering	DS Engineering	DP Engineering
	HPC 1	**DS 1**	**DP 1**
		DS 2	**DP 2**
		DS 3	**DP 3**
		DS Demo 1	*DP Demo 1*
		DS Demo 2	*DP Demo 2*
		DS Engineering	DP Engineering

Facility 1

Facility 2

Prospective Comparability Study Design

Historical (Facility 1) batches:

- Seven (7) batches
- Mix of full-scale batches:
  - Demo (non-GMP, lab)
  - Shakedown (non-GMP, mfg)
  - Engineering and Clinical (GMP, mfg)

Comparability (Facility 2) batches:

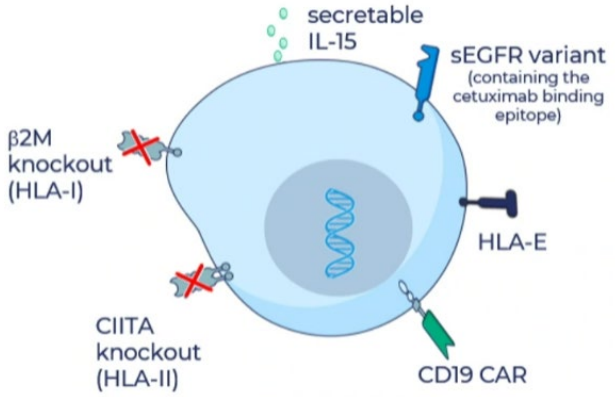
- Three (3) batches
- Matched starting material to Facility 1 (MCB1, HPC1)
- Mix of half and full-scale batches:
  - Demo (non-GMP, lab, half-scale)
  - Engineering (GMP, mfg, full-scale)

Italics: Lab-generated; Non-Italics: Manufacturing-generated; Bold: Clinical
Note: Half-scale is not a scale-down, rather half the number of parallel processing trains

Multiple orthogonal methods selected to evaluate structure and function

Structure

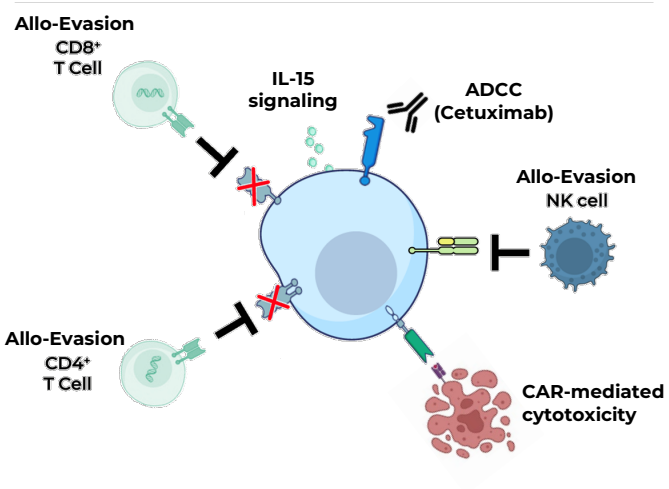
What attributes are required for a product to affect a certain function?



- Gene knockout
- Transgene on-target insertion
- Transgene sequence
- Protein expression
- Cell phenotype

Function

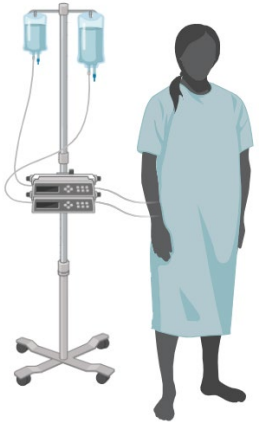
What functions (i.e., mechanisms of action) are required for biological effect?



- CAR-mediated cytotoxicity
- NK cell persistence
- Allo-evasion
- ADCC

Potency

What amount of a product (i.e., strength) is required to produce an effect?



- Dose (amount, function)
- Extrinsic factors (tumor burden, distribution, antigen expression, microenvironment)

Based on scope of the change, a mix of in-process metrics, release, and characterization data also considered

In-Process Metrics

Drug Substance Process
- Seeding density
- Cells harvested/vessel
- Lactate

Drug Product Process
- Formulation hold time
- Pre-cryo cell count
- Pre-cryo viability
- Visual inspection pass rate*

Release (select)

Identity
- Donor identity
- Engineering identity
- Cell identity

Purity / Function
- Transgene expression (flow)
- Cell phenotype (flow)

Quantity
- Cell count and viability

Impurities / Safety
- Residual iPSCs
- Karyotype
- Genome variants
- Microbial and viral testing*

Structure / Function Relationship

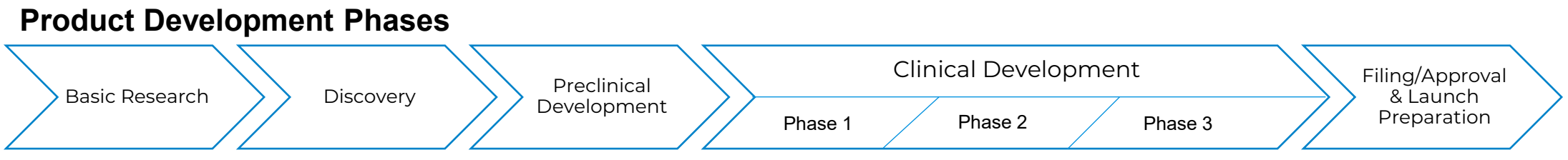
Characterization

Purity / Function
- Extended cell phenotype (flow)
- Cytotoxicity
- Degranulation
- IFN-gamma
- ADCC

Quantity
- Cell health / apoptosis

Not included for comparability:
- HPC in-process metrics, release, and stability (upstream of change)
- Some testing (*) not performed on comparability demo batches (NA – lab produced)
- DP stability not performed for comparability (but included 1 batch for annual stability)

Limited historical data during early development considered when establishing comparability criteria



Early development batches

- **Frequent process changes** make establishing representative dataset difficult

Clinical batches

- **Clinical batches** establish link to clinical performance
- **Limited manufacturing campaigns** underrepresent process and analytical variability

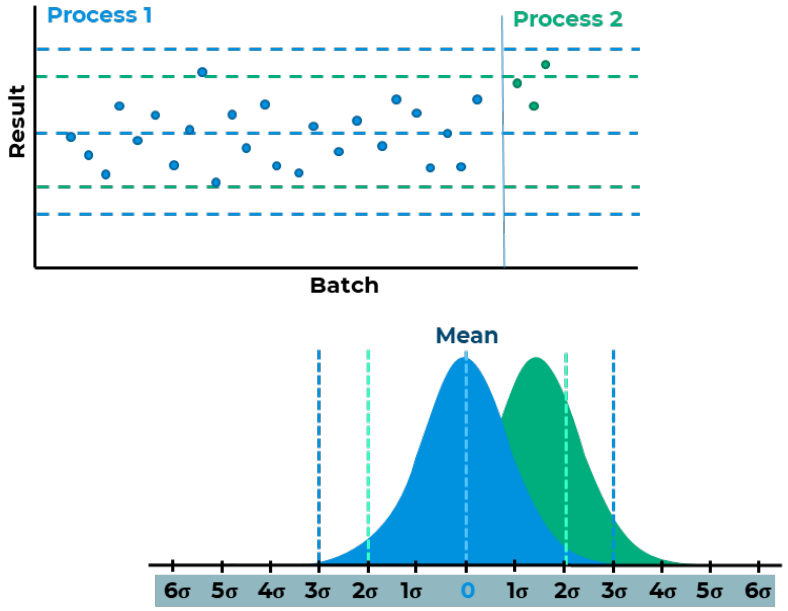
Commercial batches

- **Frequent manufacturing** provides larger dataset
- **Process drift** may occur over time

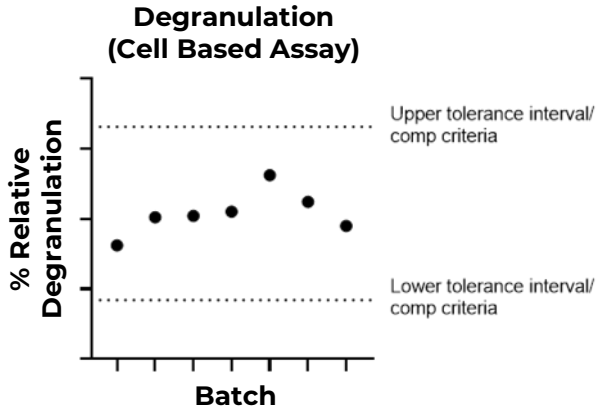
Comparability criteria include tighter alert levels in addition to specifications

Specifications
- Evaluates safety & efficacy (lot disposition)
- Based on technical justification or statistics (e.g., 3σ or 99/99 tolerance)

Comparability criteria
- Detects process shifts (investigation)
- Based on technical justification or statistics (e.g., 2σ or 95/95 tolerance)

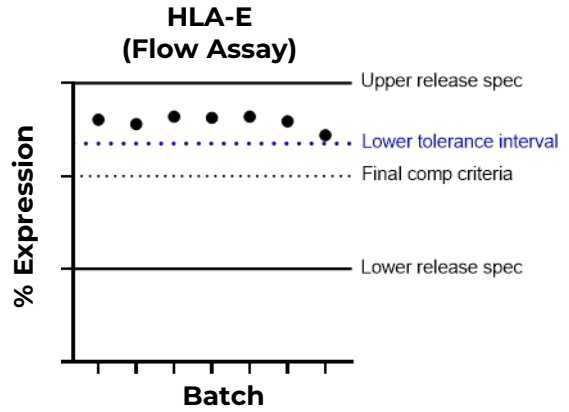


Example 1: Statistical basis



Comparability criteria set based on 95/95 tolerance limit (n=7)

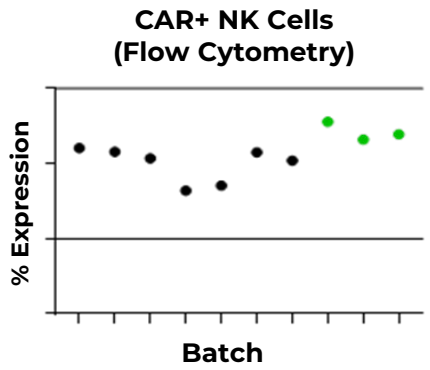
Example 2: Technical basis



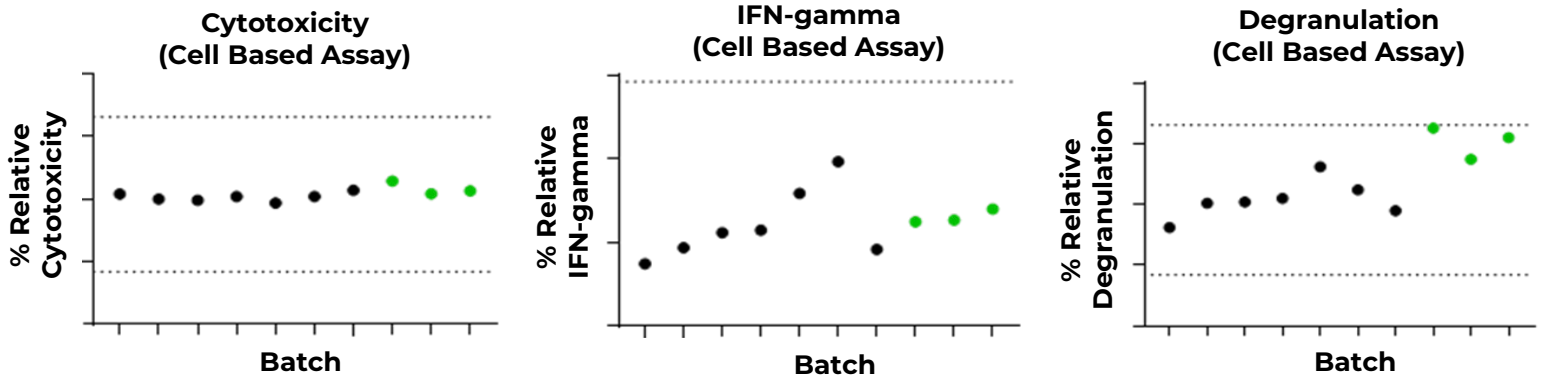
Comparability criteria set at 50% specification range based on technical justification

Assessment includes evaluation against specifications, alerts, and for shifts & trends

Release Assay



Orthogonal Characterization Assays

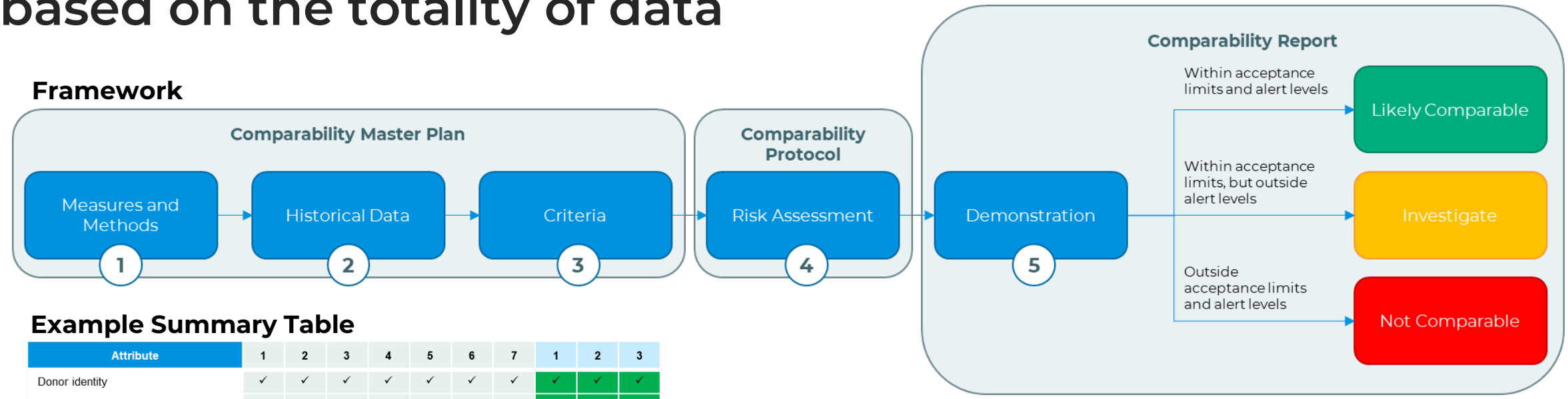


Comparability batch results (green):

- Within specifications (solid lines)
- Within comparability alerts (dashed lines)
- Potential trend noted (degranulation):
  - No observed trends in orthogonal assays (CAR+NK Cells, Cytotoxicity, IFN-gamma)
  - Change in analytical reagents may explain observed shift (control sample – not shown – also trended higher)

Comparability conclusions are based on the totality of data

Framework



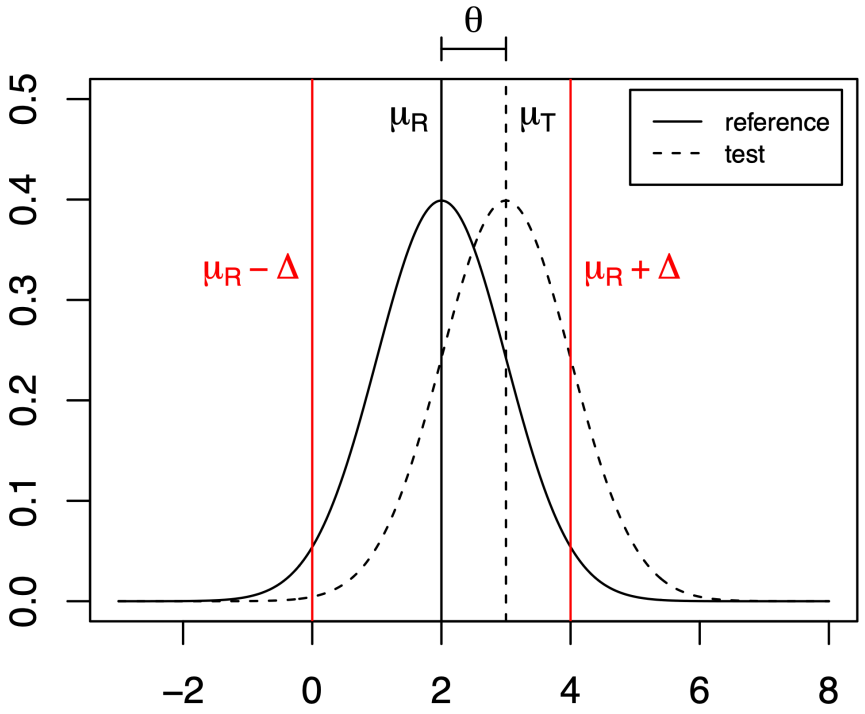
Example Summary Table

Attribute	1	2	3	4	5	6	7	1	2	3
Donor identity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Engineering identity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cell identity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Transgene expression (flow)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cell phenotype (flow)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Extended cell phenotype (flow)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cytotoxicity	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Degranulation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IFN-gamma	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cell count and viability	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Cell health and apoptosis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Residual iPSC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Karyotype	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Genome variants	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Microbial and viral safety	✓	✓	✓	✓	✓	✓	✓			✓

Facility 2 batches deemed comparable to historical Facility 1 batches:

- Comparability results generated for all attributes with evaluation against historical data, specifications, and alerts
- Conclusion based on totality of results
  - All results within specification and comparability alerts
  - Trend in one attribute identified, but not supported by orthogonal data
  - No impact to quality, safety, and efficacy
  - Non-clinical or clinical studies not required for this study

FDA recommends alternative approach for comparability of cell & gene therapy products (TOST)



- Equivalence margin (Δ)
- Reference mean (μ_R)
- Test mean (μ_T)
- Difference (Θ) between test and reference means (μ_T - μ_R)

Two one-sided t-test (TOST) determines if the mean of a test population is within some interval around the mean of a reference population:

- **Equivalence margin (Δ)** defines acceptable difference between reference and test means
- **Two t tests** define a compound pair of hypotheses:

$$H_{null} : \theta \leq -\Delta \text{ or } \Delta \leq \theta$$

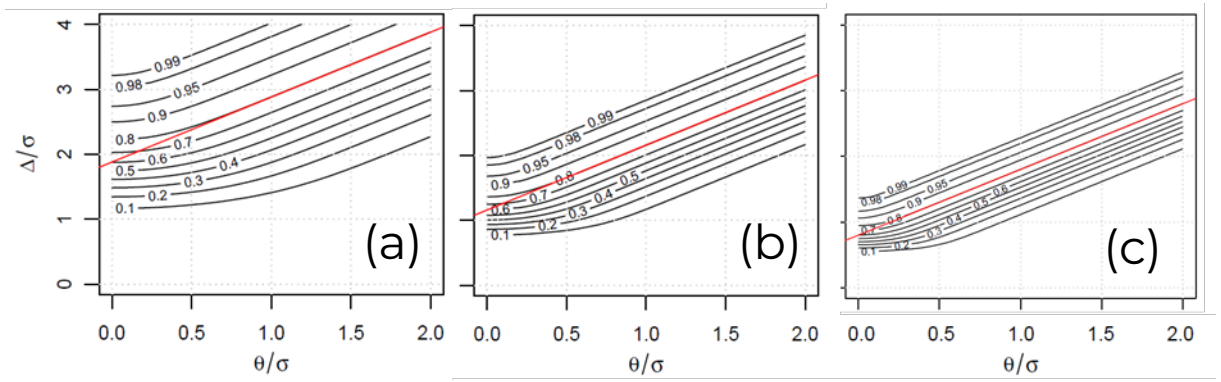
$$H_{alt} : -\Delta \leq \theta \leq \Delta$$

- **Power** - Unlike a single t test for a difference in means, TOST will only establish equivalence when it has good statistical power

Courtesy of Kevin Bullaughey

Theoretical constraints on setting equivalence margins to ensure sufficiently powered calculations

Power as a function of Δ/σ and θ/σ



Example	n_R	n_T	Δ/σ at 80% Power and θ = 0 (y-intercept)
a	7	3	1.880
b	10	10	1.160
c	20	20	0.801

- To have a high-powered test, acceptable difference (Δ) must be:
  - Larger than the true difference (θ)
  - At minimum, equal to the standard deviation (σ)
  - Large enough to account for reference and test sample size (uncertainty in μ_R, μ_T, σ)

For case study (i.e., example A), to achieve 80% power:

$$\Delta > \theta + 1.880\,\sigma$$

Courtesy of Kevin Bullaughey

Case study comparing tolerance interval and TOST approach shows similar results

Attribute	Specification	Comparability Alert (CA)	Equivalence Margin (EM)	Test Values in CA?	Signific. (TOST)
1	50.0 ± 10.0	50.0 ± 10.0	47.0 – 55.1	yes	no (low)
2	≥70.0	≥75.0	78.6 – 96.0	yes	no (high)
3	90.0 ± 10.0	90.0 ± 10.0	81.2 – 103	yes	yes
4	≥80.0	≥90.0	91.2 – 106*	yes	yes
5	≥80.0	≥90.0	89.7 – 99.2*	yes	yes
6	N/A	42.0 – 165.6	64.6 - 143	yes (high trend)	no (high)
7	N/A	42.0 – 165.6	85.2 - 118	yes	yes
8	N/A	0 – 292.9	27.9 - 213	yes	yes
9	N/A	0.36 – 5.94	1.53 – 4.77	yes (low trend)	no (low)
10	N/A	24.4 – 57.4	26.3 – 52.8	yes	yes

EM based on n=7 (reference) and n=3 (test); θ = 0.05µR (release) or 0.10µR (characterization); Power = 0.80; Equivalence Margin (Δ) > θ + 1.880 σ

* one-sided test, only consider lower bound of specification

- **Both approaches can be used to assess comparability**
  - Comparability alerts compare test values to tolerance interval or technical alerts + visual trend assessment
  - TOST compares test means to equivalence margins
- **Similar outcomes may be achievable with both assessments**
  - Attribute 1 and 2 – known shift associated with instrument change
    - Supported by assay bridging study
  - Investigation of results for attributes 6 and 9 triggered
    - Positive control tested in same run shows trend consistent with test samples
    - Results attributed to analytical variability
  - Comparability concluded

Courtesy of Kevin Bullaughey

Considerations for using comparability alerts or TOST in comparability assessments

- **During early development, limited reference and test batch sample sizes (n) may present challenges for sufficiently powered TOST assessments, due to uncertainty in μ_R, μ_T, σ**
- **If considering TOST:**
  - Increasing **reference and/or test sample size (n)** may not be possible due to:
    - High cost to manufacture larger scale, off-the-shelf batches
    - Long time to generate batches due to differentiation process
  - Widening **equivalence margins** may lead to exceeding specifications
  - Lowering acceptable **true difference (θ)** to keep EM within specification increases Type 1 error
  - Reducing **power** to allow for a more acceptable true difference increases Type 2 error

Recommendation: Prospectively discuss with agency options for early development comparability assessments