ENHANCING QSP MODEL CREATION AND VALIDATION WITH CYTOCON DB: A STANDARDIZED REPOSITORY FOR IN VIVO CELL AND CYTOKINE BASELINE CONCENTRATIONS

Oleg Demin (demin@insysbio.com),
Ekaterina Mogilevskaya, Vladislav Leonov,
Elita Gerasimuk, Nikolay Pervushin, Nail Gizzatkulov

InSysBio CY Ltd
InSysBio UK Ltd

INTRO

Calibrating Quantitative Systems Pharmacology (QSP) models requires reliable comparisons between simulations and in vivo baseline data, such as cytokine and cell concentrations from patient tissues. However, researchers face significant challenges when integrating such data from scientific literature. These include variability in measurement units (e.g., mg/mL vs. cells/mm^2), inconsistent patient demographics (e.g., age, disease stage), and diverse statistical reporting (e.g., medians, ranges). To overcome these barriers, we developed CYTOCON DB.

The primary objectives are to (1) unify fragmented literature data into a structured, queryable format, (2) enable direct comparison of model outputs with standardized biological measurements, and (3) reduce manual effort in data extraction and standardization.

METHODS

CYTOCON DB (Cell and YTOkine CONcentrations Database) is an interactive web-based platform built on the ASP.NET MVC framework, utilizing Microsoft IIS, Microsoft SQL Server, Telerik Kendo UI, and Bootstrap for a robust and user-friendly experience.

CYTOCON is a web application built on the ASP.NET MVC framework, supported by Microsoft SQL Server for data storage, and enhanced with Telerik Kendo UI and Bootstrap for user interface design. Key methodological innovations include:

1. **Unit Standardization:** Cytokine concentrations (e.g., mg/mL) are converted to pM, and cell counts (e.g., cells/mm^2) to kcell/L, using predefined formulas embedded in the database.
2. **Data Consistency & Transformation:** All types of reported averages and dispersion (e.g., median, interquartile range) are normalized to mean and standard deviation (SD). Additionally, weighted mean and SD calculations enable more accurate comparisons.
3. **Quality Control:** A two-step workflow involves annotators extracting data from literature and reviewers validating entries. Automated outlier detection algorithms help identify potential misprints or inconsistencies in source data and reduce human error.
4. **Dynamic Updates:** The database evolves continuously, with more than 1,000 new concentration values added monthly from peer-reviewed studies.
5. **API Integration:** Direct access via R and Python scripts for querying data, generating plots, and integrating results into modeling workflows.

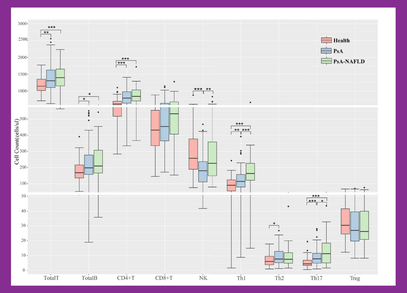
RESULTS

Database now aggregate 101,373 concentration values for cytokines and cells, curated from 2,691 scientific papers and public sources. To illustrate the capabilities of CYTOCON DB, we present two use cases:

1. IL-1α, IL-1β, IL-6, and TNF-α concentrations in serum samples from healthy controls (HC) and SLE patients were retrieved, visualized, and statistically compared. Although mean concentrations of IL-1α, IL-1β, and IL-6 were not significantly different between groups, TNF-α levels were notably higher in SLE (1.972±4.418 pM) than in HC (0.928±3.305 pM). Figure 1

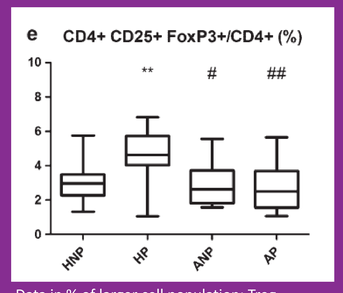
2. Cell abundance data—often reported as percentages of a parental subset (e.g., PD1+ CD8+ CD3+ cells relative to CD8+ CD3+ cells)—were transformed into absolute counts by leveraging averaged values from similar entries in the database. When applied to melanoma and NSCLC data, recalculated absolute counts (kcells/L) revealed fewer myeloid dendritic cells and natural killer cells in melanoma tumors compared to NSCLC (e.g., mDC: 1.5×10^6±2.9×10^6 vs. 1.1×10^7±4.6×10^6). Figure 2

Cell and c**YTO**kine **CON**centrations **D**ata**B**ase is a standardized, user-friendly framework for integrating, transforming, and comparing baseline concentrations across diverse data sources

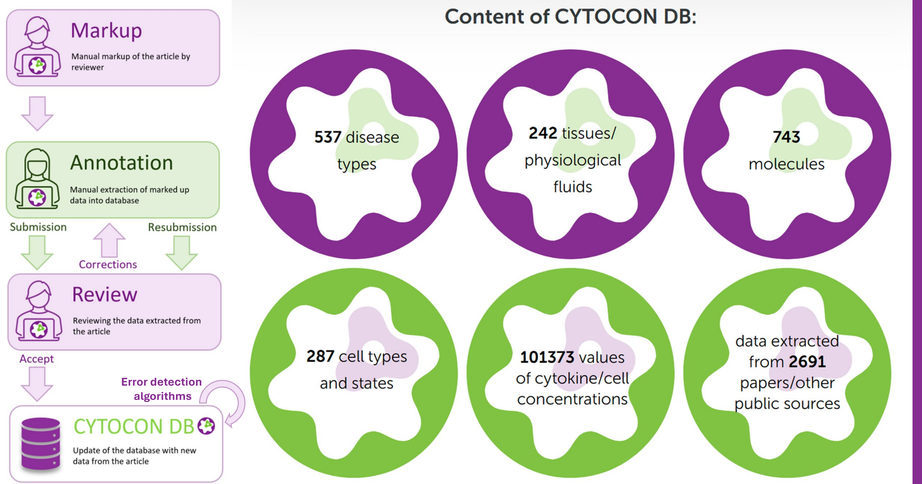


Graphical representation of the data: digitalization is required. Axes disruption may lead to errors in case of automatic digitalization.

1. Digitalization
2. Recalculation
3. Unification
4. Validation



Data in % of larger cell population: Treg (CD25+, CD4+, FoxP3+) measured as % of CD4+ T cells. If information from larger missing require data from other source.



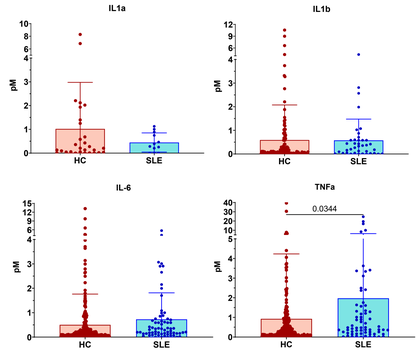
Scan QR to access **CYTOCON DB**

Scan QR to download the paper about **CYTOCON DB**

2025

CYTOCON Database

A



B

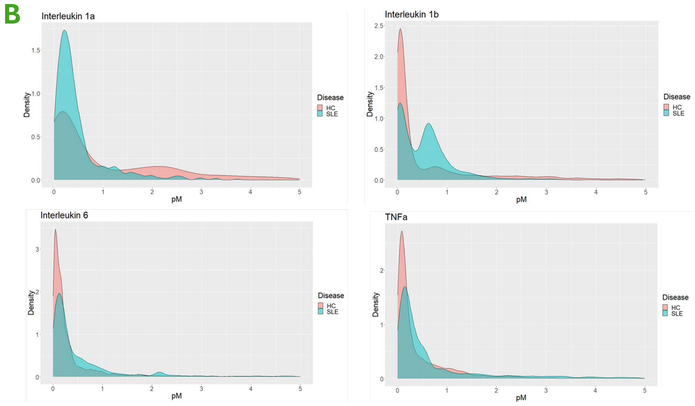


Figure 1. Bar- and Density plot in HC and SLE patients of IL1A, IL1b, IL6 and TNFa

A. Each dot represent independent source. HC – Healthy Control (IL1a: S=24, N=2173; IL1b: S=115, N=8085; IL-6: S=242, N=16245; TNFa: S=219, N=13722). SLE - Systemic Lupus Erythematosus (IL1a: S=7, N=824; IL1b: S=25, N=2037; IL-6: S=46, N=3638; TNFa: S=44, N=3399). S – source number, N – total patient number. Data represented as Mean±SD. The p-value was calculated using a t-test.
B. IL1a, IL1b, IL6, TNFa density plots are created based on the number of patients from each source, assuming normal data distribution and omitting negative values.

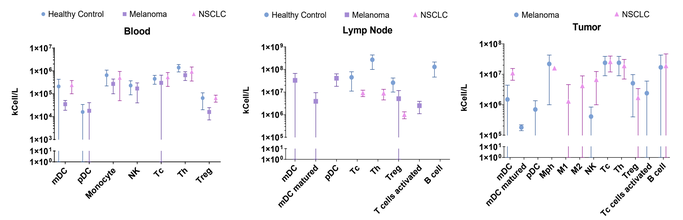


Figure 2. Comparison of unified cell/cytokine concentrations of melanoma/NSCLC patients and healthy subjects

Amount of myeloid dendritic cells and natural killer cells lower in melanoma tumors than in NSCLC (mDC 1.5E+06±2.9E+06 vs. 1.1E+07±4.6E+06; NK: 4.1E+05±4.3E+05 vs. 6.7E+06±5.7E+06). Data represented as recalculated Mean±SD with kcells/L dimension. Each dot illustrates absolute cell count value calculated as weighted average across different papers annotated in CYTOCON DB.