**Methodology: Constructing Bio-realistic Human Single-Neuron Models**

Our "bio-realistic" neuron models are designed with these key characteristics:

* Morphological Accuracy: Each model incorporates dendrite and soma morphologies derived from actual human specimen data, focusing on realistic replication of neuronal structure.
* Comprehensive Ion Channel Simulation: Unlike perisomatic models that limit active channels to the soma, our models include active sodium and potassium ion channels distributed across the dendrites, enabling more detailed simulation of neuronal activity.
* Electrophysiological Trace Matching: The models are optimized to replicate experimental electrophysiological (ephys) traces of the exact neurons for the morphology and the cell type sequencing, ensuring high fidelity to recorded data.
* Cell Type and Population Coverage: We include a range of major human neuron cell types, each validated through single-nucleus sequencing. For each cell type, a sample population represents the inherent variability within the type (see model cell count table for details).

The model structure itself is designed to balance anatomical accuracy with computational efficiency. Each morphological component is simplified into short cylindrical segments that represent parts of the dendrites and soma. On the surface of each segment, we incorporate membrane capacitance, passive leak current resistance, and voltage-gated ion channels according to the Hodgkin–Huxley model. To simulate intracellular connectivity, adjacent cylinders are joined through axial resistance, which replicates the natural spread of electrical signals within the cell.

To ensure an accurate fit to experimental performance, we defined a set of adjustable parameters for each model. These include membrane capacitance (Cm), which reflects the membrane’s thickness and total surface area; axial resistance (Rm), representing cytoplasmic variations; ion channel conductance (g\_bar), accounting for differences in ion channel density across cells; and leak current potential and conductance (e\_pas and g\_pas) to modulate passive current. We also altered two specific calcium channel properties (gamma and decay for Ca\_dynamic), which further refine the model’s fidelity to recorded neuron behavior.

Given the complexity of the model, parameter optimization is treated as a multi-parameter problem. The high degree of non-linearity led us to employ genetic optimization to identify the best-fit parameter set, with an objective function designed as the sum of deviations between model features and actual ephys data. Missing features in certain traces—such as spike counts in negative voltage traces—were omitted to focus only on applicable attributes. For normalization, we compiled datasets with multiple experimental traces, each collected using consistent clamping currents. From these, we calculated the standard deviation of each feature across traces and used it as a normalization denominator, adding further accuracy by weighting passive features four times higher to better capture voltage envelope characteristics.

We adopted a two-stage optimization strategy, as described by Nandi et al. In the first stage, we used one positive and five negative clamping currents, aiming to capture passive responses and the sag characteristic seen in response to negative input. In the second stage, we chose six traces for each cell, adapting each set to the cell's specific clamping responses. This included one trace with a strongly negative current (around -0.1 pA), another trace just before spike initiation, three traces with progressively higher currents that trigger spiking, and one trace with a strong positive current (typically around +0.2 pA or above). This second stage is designed for comprehensive coverage of each neuron’s frequency-current (f-I) response curve. (Some cells may have slightly fewer traces selected due to the experiment set differences.)

The optimization process is computationally intensive, with the first stage consisting of 50 generations of 4,096 individuals each, and the second stage extending to 200 generations with the same population size. This structure requires over 6 million simulations for a single optimization run, an effort aimed at thoroughly exploring the parameter space for an optimal fit. Each cell model typically demands approximately 5,000 CPU-hours to complete this optimization process. This estimate is based on the use of high-performance, compute-intensive CPUs available in 2023, such as Intel Xeon or AMD Epyc processors, which are capable of handling the high-frequency, parallelized computations necessary for efficient simulation. We performed the optimization using different randomization seeds, two seeds for the excitatory cells and three for the inhibitory cells, which double or triple the compute cost. (See cell population table)

As a final quality control (QC) measure, we evaluated each model based on two key electrophysiological features: "first afterhyperpolarization (AHP) depth from peak" and "first spike width." These features serve as indicators of the model's ability to replicate the nuanced aspects of real neuronal behavior. Any model falling outside the ±40% range of the experimental values for these features was disqualified. From each cell's optimization results, we selected the top 40 performers based on their fit to the target data. If, after applying the QC criteria, fewer than 20 models remained within the acceptable range, we disqualified the entire batch. Otherwise, we retained the top 20 models, which we documented as the final set of validated, high-quality models for each cell type. This process ensured that only the most accurate and reliable models were included in the final dataset.

Table 1 Model Cell Population

(x20 usable model each)

'ITL23' 'ITL35' 'ITL46' 'ITL6' 'L56NP'

83 40 24 1 2

82 40 24 1 2

'LAMP5PAX6Other' 'PVALB' 'SST' 'VIP'

40 18 7 15

43 17 7 16

41 17 7 15

Table 2 model mechanism used (other than “pas”)

Inhibitory neurons

"dendrites":

"Ih",

"NaV",

"Kv3\_1",

"Im\_v2"

"somatic":

"Ih",

"Kv3\_1",

"NaV",

"K\_Tst",

"K\_Pst",

"SK",

"Ca\_HVA",

"Ca\_LVA",

"CaDynamics"

Excitatory neurons

"dendrites":

"Ih",

"NaTs2\_t",

"Im",

"Kv3\_1"

"somatic":

"Ih",

"Kv3\_1",

"NaTs2\_t",

"Nap\_Et2",

"K\_Tst",

"K\_Pst",

"SK",

"Ca\_HVA",

"Ca\_LVA",

"CaDynamics"

Table 3: Feature used in optimization

Stage 1

decay\_time\_constant\_after\_stim

sag\_amplitude

steady\_state\_voltage

steady\_state\_voltage\_stimend

voltage\_base

Stage 2

AHP1\_depth\_from\_peak

AHP\_depth

AHP\_time\_from\_peak

AP1\_peak

AP1\_width (x4)

Spikecount

decay\_time\_constant\_after\_stim

depol\_block

inv\_first\_ISI

mean\_AP\_amplitude

sag\_amplitude

steady\_state\_voltage (x4)

steady\_state\_voltage\_stimend

time\_to\_first\_spike

Voltage\_base (x4)