Supplementary Information

Machine learning to predict developmental neurotoxicity with high-throughput data from 2D bio-engineered tissues

Finn Kuusisto Vitor Santos Costa Zhonggang Hou James Thomson
David Page Ron Stewart

List of supplementary content

Supplementary Text S1	Cell culture and neural differentiation
Supplementary Text S2	RNA sequencing
Supplementary Table T1	Compounds and concentrations used

Supplementary Text S1. Cell culture and neural differentiation

For our 2D tissue model, we maintained human H1 embryonic stem (ES) cells in E8 medium [1] on Matrigel (BD Biosciences) coated plates and passaged every 4-5 days with EDTA [2]. To derive neural progenitor cells (NPC), we split H1 cells using EDTA and immediately treated them with E7 medium (E8 medium without TGFβ-1) +SB431542 (Sigma) (10μM). After another seven days, we mechanically dissociated the formed rosettes from the culture dish and cultured the floating aggregates in E5 medium (E8 medium without FGF2, TGFβ-1, or insulin) +N2+B27+hFGF2 (4ng/ml). Four days later, we dissociated the aggregates with Accutase (Life Technologies Inc.) and plated them onto Matrigel coated plates in E5+N2+B27+hFGF2 (5ng/ml). We cultured the cells for an additional 22 days and passaged when confluent. After culture, we cryopreserved the cells at 12 million cells per vial. We performed fluorescence-activated cell sorting (FACS) analysis and found that these NPCs were greater than 90% SOX1$^+$ and TubβIII$^+$.

We thawed the NPCs and expanded them in E5+N2+B27+hFGF2 (5ng/ml) for five days before harvesting via Accutase treatment. We then seeded roughly 10^5 cells into one well of a 48 well Matrigel coated plate in E5+N2+B27 and started chemical treatment the same day (day 0). We collected samples at indicated time points by lysing cells directly on plate with 150μl RLT lysis buffer (Qiagen).

Supplementary Text S2. RNA sequencing

After treatment with the compound, we purified total RNA from the NPCs using the RNeasy 96 Kit (Qiagen). We prepared cDNA libraries, indexed with Illumina's TruSeq RNA Sequencing Kit, and sequenced on Illumina's HiSeq 2500, with 17-24 indexed samples per lane with 51 base-pair single reads. We generated FASTQ files with CASAVA (v1.8.2) and mapped reads to the human transcriptome (RefGene v1.1.17) using Bowtie [3] (v0.12.8), allowing 2 mismatches and a maximum of 20 multiple hits. To produce gene expression values in TPM, we used RSEM [4] (v1.2.1).

Supplementary Table T1. Compounds and concentrations used

	Chemical	Concentration
Toxic	Accutane or Isotretinoin	0.5μg/ml
	Amiodarone	1μg/ml
	Arsenic	5μM
	Benzene	50μg/ml
	Bioallethrin	10nM
	Bis-I	4μM
	Busulfan	50μg/ml
	Cadmium	50μM
	Carbamazepine	10μg/ml
	Cytosine β-D-arabinofuranoside	5μM
	Dexamethasone	100μM
	Diazinon	10μM
	Dioxin	3nM
	5-Fluorouracil	10μM
	Hydroxyurea	250μM
	2-Imidazolidinethione	0.4mM
	K252a	30nM
	Lead acetate	30μM
	Maneb	60μM
	Monosodium glutamate	150μM
	Okadaic acid	3nM
	PD166866	2.5μM
	Permethrin	2.5μM
	L-phenylalanine	1mM
	Propylthiouracil	25μg/ml
	(trans) Retinoic acid	1.7μM
	Thalidomide	100ng/ml
	U0126	10μM
	Vincristine	1nM
Non-Toxic	Acetaminophen	10μg/ml
	Amoxicillin	5μg/ml
	Aspirin	20μg/ml
	DMSO	0.1%
	Ficoll 400	10μM
	Fructose	10μM
	Glucosamine	0.4μg/ml
	Glycerol	10μM
	Glyphosate	10μM
	Ibuprofen	20μg/ml
	Naproxen sodium	30μg/ml
	PEG 3350	10μM
	PVP	100μg/ml
	Saccharin	10μM
	Sorbitol	10μM
	Sucrose	10μM

References

[1] G. Chen, D. R. Gulbranson, Z. Hou, J. M. Bolin, V. Ruotti, M. D. Probasco, K. Smuga-Otto, S. E. Howden, N. R. Diol, N. E. Propson *et al.*, "Chemically defined conditions for human ipsc derivation and culture," *Nature Methods*, vol. 8, no. 5, pp. 424–429, 2011.

[2] J. Beers, D. R. Gulbranson, N. George, L. I. Siniscalchi, J. Jones, J. A. Thomson, and G. Chen, "Passaging and colony expansion of human pluripotent stem cells by enzyme-free dissociation in chemically defined culture conditions," *Nature Protocols*, vol. 7, no. 11, pp. 2029–2040, 2012.

[3] B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg, "Ultrafast and memory-efficient alignment of short dna sequences to the human genome," *Genome Biology*, vol. 10, no. 3, p. R25, 2009.

[4] B. Li and C. N. Dewey, "RSEM: Accurate transcript quantification from rna-seq data with or without a reference genome," *BMC Bioinformatics*, vol. 12, no. 1, p. 323, 2011.