

## Supplementary Tables

### CrossTx: Cross-cell line Transcriptomic Signature Predictions

Panagiotis Chrysinas<sup>1</sup>, Changyou Chen<sup>2</sup>, Rudiyanto Gunawan<sup>1,\*</sup>

<sup>1</sup>Department of Chemical and Biological Engineering, University at Buffalo-SUNY, Buffalo, NY 14260, USA

<sup>2</sup>Department of Computer Science and Engineering, University at Buffalo-SUNY, Buffalo, NY 14260, USA

**Supplementary Table S1. Pearson correlations of drug signature predictions by CrossTx and TT-WOPT for individual cell lines.** Values correspond to mean  $\pm$  standard deviation. Bold values signify the best method based on Pearson correlation. Statistical significance was established by two-sided paired t-test to assessed the change in accuracy by adding a Corrector. For example, Mean + PCA was compared to Mean, while Mean + PCA + AE was compared to Mean + PCA. Note that the addition of a Corrector may degrade accuracy. TT-WOPT was compared to the Mean method. \*:  $p$ -values < 0.05.

Cell type \ Method	MCF7	A375	HT29	PC3	HA1E	YAPC	HELA
<b>Predictor: Mean (<math>\mu</math>) + Corrector: PCA, AE, PCA + AE, AE + PCA</b>							
Mean	0.59 $\pm 0.14$	0.58 $\pm 0.12$	0.6 $\pm 0.13$	0.59 $\pm 0.14$	0.58 $\pm 0.14$	0.58 $\pm 0.18$	0.62 $\pm 0.11$
Mean + PCA	<b>0.76*</b> $\pm 0.17$	<b>0.77*</b> $\pm 0.15$	<b>0.77*</b> $\pm 0.18$	<b>0.75*</b> $\pm 0.19$	<b>0.74*</b> $\pm 0.21$	<b>0.72*</b> $\pm 0.23$	<b>0.76*</b> $\pm 0.2$
Mean + AE	<b>0.72*</b> $\pm 0.15$	<b>0.67*</b> $\pm 0.15$	<b>0.73*</b> $\pm 0.14$	<b>0.71*</b> $\pm 0.16$	<b>0.67*</b> $\pm 0.16$	<b>0.69*</b> $\pm 0.22$	<b>0.76*</b> $\pm 0.12$
Mean + PCA + AE	<b>0.78*</b> $\pm 0.17$	<b>0.78*</b> $\pm 0.16$	<b>0.78*</b> $\pm 0.18$	<b>0.77*</b> $\pm 0.19$	<b>0.75*</b> $\pm 0.22$	<b>0.74*</b> $\pm 0.24$	<b>0.79*</b> $\pm 0.19$
Mean + AE + PCA	<b>0.75*</b> $\pm 0.18$	<b>0.76*</b> $\pm 0.16$	<b>0.76*</b> $\pm 0.18$	<b>0.73*</b> $\pm 0.21$	<b>0.73*</b> $\pm 0.22$	0.69 $\pm 0.25$	0.76 $\pm 0.2$
<b>Predictor: Regression + Corrector: PCA, AE, PCA + AE, AE + PCA</b>							
Regression	0.55 $\pm 0.15$	0.56 $\pm 0.14$	0.56 $\pm 0.17$	0.56 $\pm 0.16$	0.54 $\pm 0.16$	0.52 $\pm 0.22$	0.54 $\pm 0.16$
Regression + PCA	<b>0.62*</b> $\pm 0.25$	<b>0.69*</b> $\pm 0.22$	<b>0.66*</b> $\pm 0.25$	<b>0.61*</b> $\pm 0.26$	<b>0.6*</b> $\pm 0.27$	<b>0.62*</b> $\pm 0.27$	<b>0.63*</b> $\pm 0.27$
Regression + AE	<b>0.46*</b> $\pm 0.26$	<b>0.25*</b> $\pm 0.38$	<b>0.44*</b> $\pm 0.3$	<b>0.35*</b> $\pm 0.34$	<b>0.23*</b> $\pm 0.41$	<b>0.38*</b> $\pm 0.31$	<b>0.49*</b> $\pm 0.29$
Regression + PCA + AE	<b>0.66*</b> $\pm 0.24$	<b>0.72*</b> $\pm 0.22$	<b>0.68*</b> $\pm 0.26$	<b>0.63*</b> $\pm 0.26$	<b>0.61*</b> $\pm 0.29$	<b>0.66*</b> $\pm 0.26$	<b>0.66*</b> $\pm 0.26$
Regression + AE + PCA	0.45 $\pm 0.3$	<b>0.28*</b> $\pm 0.47$	<b>0.55*</b> $\pm 0.31$	<b>0.39*</b> $\pm 0.35$	0.23 $\pm 0.46$	<b>0.47*</b> $\pm 0.29$	0.47 $\pm 0.31$
<b>TT-WOPT</b>							
TT-WOPT	<b>0.31*</b> $\pm 0.16$	<b>0.19*</b> $\pm 0.2$	<b>0.18*</b> $\pm 0.18$	<b>0.05*</b> $\pm 0.12$	<b>0.36*</b> $\pm 0.15$	<b>0.18*</b> $\pm 0.15$	<b>0.24*</b> $\pm 0.2$

**Supplementary Table S2. AUPRs of drug signature predictions by CrossTx using the Mean method as Predictor and TT-WOPT for individual cell lines.** Values correspond to mean  $\pm$  standard deviation. Bold values signify the best method based on Pearson correlation. Statistical significance was established by two-sided paired t-test to assessed the change in accuracy by adding a Corrector. For example, Mean + PCA was compared to Mean, while Mean + PCA + AE was compared to Mean + PCA. Note that the addition of a Corrector may degrade accuracy. TT-WOPT was compared to the Mean method. \*:  $p$ -values < 0.05.

Method \ Cell type	MCF7	A375	HT29	PC3	HA1E	YAPC	HELA
<b>AUPR for predicting upregulated genes</b>							
Mean	0.66 $\pm 0.11$	0.65 $\pm 0.1$	0.66 $\pm 0.11$	0.65 $\pm 0.12$	0.65 $\pm 0.12$	0.65 $\pm 0.14$	0.68 $\pm 0.11$
Mean + PCA	<b>0.78*</b> $\pm 0.14$	<b>0.79*</b> $\pm 0.13$	<b>0.79*</b> $\pm 0.15$	<b>0.77*</b> $\pm 0.15$	<b>0.77*</b> $\pm 0.17$	<b>0.75*</b> $\pm 0.17$	<b>0.79*</b> $\pm 0.16$
Mean + AE	<b>0.75*</b> $\pm 0.13$	<b>0.73*</b> $\pm 0.13$	<b>0.76*</b> $\pm 0.14$	<b>0.74*</b> $\pm 0.14$	<b>0.72*</b> $\pm 0.14$	<b>0.73*</b> $\pm 0.17$	<b>0.78*</b> $\pm 0.12$
Mean + PCA + AE	<b>0.79*</b> $\pm 0.14$	<b>0.80*</b> $\pm 0.13$	<b>0.80*</b> $\pm 0.15$	<b>0.79*</b> $\pm 0.16$	<b>0.78*</b> $\pm 0.17$	<b>0.78*</b> $\pm 0.17$	<b>0.81*</b> $\pm 0.15$
Mean + AE + PCA	<b>0.77*</b> $\pm 0.14$	<b>0.78*</b> $\pm 0.13$	<b>0.78*</b> $\pm 0.15$	<b>0.76*</b> $\pm 0.17$	<b>0.76*</b> $\pm 0.17$	0.74 $\pm 0.18$	0.79 $\pm 0.15$
TT-WOPT	<b>0.48*</b> $\pm 0.1$	<b>0.431*</b> $\pm 0.095$	<b>0.428*</b> $\pm 0.095$	<b>0.347*</b> $\pm 0.087$	<b>0.52*</b> $\pm 0.12$	<b>0.405*</b> $\pm 0.097$	<b>0.44*</b> $\pm 0.13$
<b>AUPR for predicting downregulated genes</b>							
Mean	0.65 $\pm 0.11$	0.67 $\pm 0.11$	0.67 $\pm 0.11$	0.65 $\pm 0.13$	0.66 $\pm 0.13$	0.68 $\pm 0.12$	0.686 $\pm 0.096$
Mean + PCA	<b>0.79*</b> $\pm 0.14$	<b>0.82*</b> $\pm 0.13$	<b>0.81*</b> $\pm 0.14$	<b>0.78*</b> $\pm 0.16$	<b>0.79*</b> $\pm 0.17$	<b>0.79*</b> $\pm 0.15$	<b>0.81*</b> $\pm 0.15$
Mean + AE	<b>0.76*</b> $\pm 0.12$	<b>0.74*</b> $\pm 0.13$	<b>0.78*</b> $\pm 0.13$	<b>0.74*</b> $\pm 0.15$	<b>0.72*</b> $\pm 0.14$	<b>0.76*</b> $\pm 0.15$	<b>0.80*</b> $\pm 0.1$
Mean + PCA + AE	<b>0.81*</b> $\pm 0.14$	<b>0.83*</b> $\pm 0.13$	<b>0.82*</b> $\pm 0.14$	<b>0.80*</b> $\pm 0.16$	<b>0.80*</b> $\pm 0.17$	<b>0.81*</b> $\pm 0.16$	<b>0.83*</b> $\pm 0.14$
Mean + AE + PCA	<b>0.79*</b> $\pm 0.15$	<b>0.80*</b> $\pm 0.13$	<b>0.81*</b> $\pm 0.14$	<b>0.77*</b> $\pm 0.17$	<b>0.78*</b> $\pm 0.17$	0.77 $\pm 0.16$	0.80 $\pm 0.15$
TT-WOPT	<b>0.50*</b> $\pm 0.11$	<b>0.46*</b> $\pm 0.12$	<b>0.43*</b> $\pm 0.11$	<b>0.365*</b> $\pm 0.097$	<b>0.53*</b> $\pm 0.14$	<b>0.448*</b> $\pm 0.087$	<b>0.48*</b> $\pm 0.12$

**Supplementary Table S3. AUPRs of drug signature predictions by CrossTx using the Regression method as Predictor and TT-WOPT for individual cell lines.** Values correspond to mean  $\pm$  standard deviation. Bold values signify the best method based on Pearson correlation. Statistical significance was established by two-sided paired t-test to assessed the change in accuracy by adding a Corrector. For example, Mean + PCA was compared to Mean, while Mean + PCA + AE was compared to Mean + PCA. Note that the addition of a Corrector may degrade accuracy. TT-WOPT was compared to the Regression method. \*:  $p$ -values < 0.05.

Cell type \ Method	MCF7	A375	HT29	PC3	HA1E	YAPC	HELA
<b>AUPR for predicting upregulated genes</b>							
Regression	0.64 $\pm 0.12$	0.64 $\pm 0.11$	0.63 $\pm 0.13$	0.63 $\pm 0.13$	0.62 $\pm 0.14$	0.61 $\pm 0.15$	0.63 $\pm 0.13$
Regression + PCA	<b>0.67*</b> $\pm 0.18$	<b>0.74*</b> $\pm 0.16$	<b>0.71*</b> $\pm 0.19$	<b>0.66*</b> $\pm 0.2$	<b>0.66*</b> $\pm 0.19$	<b>0.68*</b> $\pm 0.2$	<b>0.69*</b> $\pm 0.19$
Regression + AE	<b>0.58*</b> $\pm 0.18$	<b>0.51*</b> $\pm 0.21$	<b>0.58*</b> $\pm 0.2$	<b>0.52*</b> $\pm 0.22$	<b>0.49*</b> $\pm 0.22$	<b>0.55*</b> $\pm 0.2$	<b>0.60*</b> $\pm 0.2$
Regression + PCA + AE	<b>0.71*</b> $\pm 0.18$	<b>0.77*</b> $\pm 0.17$	<b>0.73*</b> $\pm 0.2$	<b>0.69*</b> $\pm 0.2$	<b>0.69*</b> $\pm 0.2$	<b>0.71*</b> $\pm 0.2$	<b>0.72*</b> $\pm 0.19$
Regression + AE + PCA	0.58 $\pm 0.19$	<b>0.54*</b> $\pm 0.25$	<b>0.64*</b> $\pm 0.22$	<b>0.54*</b> $\pm 0.23$	<b>0.50*</b> $\pm 0.24$	<b>0.59*</b> $\pm 0.2$	0.59 $\pm 0.22$
TT-WOPT	<b>0.48*</b> $\pm 0.1$	<b>0.431*</b> $\pm 0.095$	<b>0.428*</b> $\pm 0.095$	<b>0.347*</b> $\pm 0.087$	<b>0.52*</b> $\pm 0.12$	<b>0.405*</b> $\pm 0.097$	<b>0.44*</b> $\pm 0.13$
<b>AUPR for predicting downregulated genes</b>							
Regression	0.63 $\pm 0.12$	0.67 $\pm 0.11$	0.65 $\pm 0.12$	0.64 $\pm 0.14$	0.64 $\pm 0.14$	0.65 $\pm 0.14$	0.64 $\pm 0.12$
Regression + PCA	<b>0.69*</b> $\pm 0.19$	<b>0.76*</b> $\pm 0.16$	<b>0.73*</b> $\pm 0.18$	<b>0.68*</b> $\pm 0.2$	<b>0.69*</b> $\pm 0.19$	<b>0.72*</b> $\pm 0.19$	<b>0.71*</b> $\pm 0.19$
Regression + AE	<b>0.59*</b> $\pm 0.18$	<b>0.5*</b> $\pm 0.22$	<b>0.58*</b> $\pm 0.21$	<b>0.52*</b> $\pm 0.22$	<b>0.48*</b> $\pm 0.22$	<b>0.56*</b> $\pm 0.21$	<b>0.63*</b> $\pm 0.19$
Regression + PCA + AE	<b>0.73*</b> $\pm 0.18$	<b>0.78*</b> $\pm 0.16$	<b>0.76*</b> $\pm 0.19$	<b>0.70*</b> $\pm 0.2$	<b>0.71*</b> $\pm 0.2$	<b>0.75*</b> $\pm 0.18$	<b>0.74*</b> $\pm 0.18$
Regression + AE + PCA	<b>0.58*</b> $\pm 0.21$	<b>0.53*</b> $\pm 0.26$	<b>0.66*</b> $\pm 0.22$	<b>0.55*</b> $\pm 0.23$	<b>0.49*</b> $\pm 0.26$	<b>0.61*</b> $\pm 0.2$	0.63 $\pm 0.2$
TT-WOPT	<b>0.50*</b> $\pm 0.11$	<b>0.46*</b> $\pm 0.12$	<b>0.43*</b> $\pm 0.11$	<b>0.365*</b> $\pm 0.097$	<b>0.53*</b> $\pm 0.14$	<b>0.448*</b> $\pm 0.087$	<b>0.48*</b> $\pm 0.12$

**Supplementary Table S4. AUROCs of drug signature predictions by CrossTx using the Mean method as Predictor and TT-WOPT for individual cell lines.** Values correspond to mean  $\pm$  standard deviation. Bold values signify the best method based on Pearson correlation. Statistical significance was established by two-sided paired t-test to assessed the change in accuracy by adding a Corrector. For example, Mean + PCA was compared to Mean, while Mean + PCA + AE was compared to Mean + PCA. Note that the addition of a Corrector may degrade accuracy. TT-WOPT was compared to the Mean method. \*:  $p$ -values  $< 0.05$ .

Cell type \ Method	MCF7	A375	HT29	PC3	HA1E	YAPC	HELA
<b>AUROC for predicting upregulated genes</b>							
Mean	0.782 $\pm 0.069$	0.779 $\pm 0.064$	0.787 $\pm 0.068$	0.791 $\pm 0.067$	0.785 $\pm 0.073$	0.781 $\pm 0.092$	0.799 $\pm 0.064$
Mean + PCA	0.87* $\pm 0.085$	0.878* $\pm 0.079$	0.874* $\pm 0.096$	0.869* $\pm 0.096$	0.86* $\pm 0.11$	0.85* $\pm 0.12$	0.87* $\pm 0.1$
Mean + AE	0.854* $\pm 0.078$	0.835* $\pm 0.082$	0.858* $\pm 0.08$	0.854* $\pm 0.08$	0.835* $\pm 0.085$	0.84* $\pm 0.12$	0.877* $\pm 0.069$
Mean + PCA + AE	<b>0.88*</b> <b><math>\pm 0.087</math></b>	<b>0.886*</b> <b><math>\pm 0.081</math></b>	<b>0.883*</b> <b><math>\pm 0.098</math></b>	<b>0.881*</b> <b><math>\pm 0.097</math></b>	<b>0.87*</b> <b><math>\pm 0.11</math></b>	<b>0.86*</b> <b><math>\pm 0.12</math></b>	<b>0.89*</b> <b><math>\pm 0.1</math></b>
Mean + AE + PCA	0.867* $\pm 0.091$	0.871* $\pm 0.083$	0.87* $\pm 0.095$	0.86 $\pm 0.11$	0.86* $\pm 0.11$	0.84 $\pm 0.13$	0.87 $\pm 0.1$
TT-WOPT	0.647* $\pm 0.08$	0.592* $\pm 0.097$	0.586* $\pm 0.09$	0.522* $\pm 0.061$	0.683* $\pm 0.08$	0.572* $\pm 0.074$	0.618* $\pm 0.099$
<b>AUROC for predicting downregulated genes</b>							
Mean	0.80 $\pm 0.07$	0.788 $\pm 0.065$	0.805 $\pm 0.065$	0.798 $\pm 0.075$	0.789 $\pm 0.07$	0.793 $\pm 0.087$	0.812 $\pm 0.055$
Mean + PCA	0.889* $\pm 0.086$	0.891* $\pm 0.076$	0.892* $\pm 0.086$	0.878* $\pm 0.098$	0.87* $\pm 0.11$	0.87* $\pm 0.11$	0.891* $\pm 0.088$
Mean + AE	0.872* $\pm 0.078$	0.841* $\pm 0.081$	0.874* $\pm 0.075$	0.857* $\pm 0.089$	0.838* $\pm 0.085$	0.86* $\pm 0.11$	0.892* $\pm 0.059$
Mean + PCA + AE	<b>0.898*</b> <b><math>\pm 0.086</math></b>	<b>0.897*</b> <b><math>\pm 0.079</math></b>	<b>0.901*</b> <b><math>\pm 0.088</math></b>	<b>0.89*</b> <b><math>\pm 0.1</math></b>	<b>0.88*</b> <b><math>\pm 0.11</math></b>	<b>0.88*</b> <b><math>\pm 0.11</math></b>	<b>0.903*</b> <b><math>\pm 0.086</math></b>
Mean + AE + PCA	0.884* $\pm 0.092$	0.883* $\pm 0.081$	0.888* $\pm 0.087$	0.87* $\pm 0.11$	0.87* $\pm 0.11$	0.86 $\pm 0.12$	0.889 $\pm 0.09$
TT-WOPT	0.686* $\pm 0.089$	0.60* $\pm 0.11$	0.605* $\pm 0.096$	0.545* $\pm 0.064$	0.694* $\pm 0.076$	0.621* $\pm 0.075$	0.643* $\pm 0.098$

**Supplementary Table S5. AUROCs of drug signature predictions by CrossTx using the Regression method as Predictor and TT-WOPT for individual cell lines.** Values correspond to mean  $\pm$  standard deviation. Bold values signify the best method based on Pearson correlation. Statistical significance was established by two-sided paired t-test to assessed the change in accuracy by adding a Corrector. For example, Mean + PCA was compared to Mean, while Mean + PCA + AE was compared to Mean + PCA. Note that the addition of a Corrector may degrade accuracy. TT-WOPT was compared to the Regression method. \*:  $p$ -values < 0.05.

Cell type \ Method	MCF7	A375	HT29	PC3	HA1E	YAPC	HELA
<b>AUROC for predicting upregulated genes</b>							
Regression	0.765 $\pm 0.078$	0.774 $\pm 0.07$	0.767 $\pm 0.089$	0.772 $\pm 0.079$	0.764 $\pm 0.086$	0.75 $\pm 0.11$	0.764 $\pm 0.088$
Regression + PCA	<b>0.80*</b> $\pm 0.12$	<b>0.84*</b> $\pm 0.11$	<b>0.82*</b> $\pm 0.13$	<b>0.80*</b> $\pm 0.13$	<b>0.80*</b> $\pm 0.13$	<b>0.80*</b> $\pm 0.14$	<b>0.81*</b> $\pm 0.14$
Regression + AE	0.73* $\pm 0.13$	0.63* $\pm 0.19$	0.72* $\pm 0.15$	0.67* $\pm 0.17$	0.62* $\pm 0.21$	0.69* $\pm 0.15$	0.74* $\pm 0.15$
Regression + PCA + AE	<b>0.82*</b> $\pm 0.12$	<b>0.86*</b> $\pm 0.11$	<b>0.83*</b> $\pm 0.14$	<b>0.81*</b> $\pm 0.13$	<b>0.8*</b> $\pm 0.15$	<b>0.82*</b> $\pm 0.14$	<b>0.82*</b> $\pm 0.14$
Regression + AE + PCA	0.73 $\pm 0.14$	0.64* $\pm 0.23$	0.76* $\pm 0.16$	0.69* $\pm 0.18$	0.62 $\pm 0.23$	0.73* $\pm 0.14$	0.73 $\pm 0.16$
TT-WOPT	0.647 $\pm 0.08$	0.592 $\pm 0.097$	0.586 $\pm 0.09$	0.522 $\pm 0.061$	0.683 $\pm 0.08$	0.572 $\pm 0.074$	0.618 $\pm 0.099$
<b>AUROC for predicting downregulated genes</b>							
Regression	0.786 $\pm 0.075$	0.784 $\pm 0.068$	0.788 $\pm 0.084$	0.787 $\pm 0.081$	0.773 $\pm 0.081$	0.77 $\pm 0.1$	0.781 $\pm 0.076$
Regression + PCA	<b>0.83*</b> $\pm 0.12$	<b>0.86*</b> $\pm 0.1$	<b>0.84*</b> $\pm 0.12$	<b>0.82*</b> $\pm 0.12$	<b>0.81*</b> $\pm 0.13$	<b>0.83*</b> $\pm 0.13$	<b>0.83*</b> $\pm 0.12$
Regression + AE	0.75* $\pm 0.13$	0.61* $\pm 0.2$	0.73* $\pm 0.15$	0.68* $\pm 0.18$	0.60* $\pm 0.22$	0.70* $\pm 0.16$	0.77* $\pm 0.13$
Regression + PCA + AE	<b>0.85*</b> $\pm 0.11$	<b>0.87*</b> $\pm 0.11$	<b>0.85*</b> $\pm 0.12$	<b>0.83*</b> $\pm 0.13$	<b>0.82*</b> $\pm 0.14$	<b>0.85*</b> $\pm 0.12$	<b>0.85*</b> $\pm 0.12$
Regression + AE + PCA	0.75 $\pm 0.15$	0.63* $\pm 0.25$	0.79* $\pm 0.15$	0.71* $\pm 0.17$	0.61 $\pm 0.24$	0.74* $\pm 0.15$	0.77 $\pm 0.14$
TT-WOPT	0.686* $\pm 0.089$	0.60* $\pm 0.11$	0.605* $\pm 0.096$	0.545* $\pm 0.064$	0.694* $\pm 0.076$	0.621* $\pm 0.075$	0.643* $\pm 0.098$