



# **Gene Expression Data Analysis to Identify Drug-Response Specific Gene Networks Using Matrix Factorization**

Prof: Park Sang Hyun  
Assistant: Ha Ji Hwan  
멜론빙수 (Melon Iceflake)

Lee Sun Ho | Nam Hyo Rim | Moon Chan Gyung

# Background

---

**Phenotype** an observable property of an organism such as rash, determined by gene response

**Gene Expression** A gene is called expressed when it produces (usually) proteins.

**GEO** Gene Expression Omnibus, a public gene data repository

**Matrix Factorization** a machine learning technique that uses factorizing of a matrix. A factor is called as "latent space".

**Normalized Cosine Similarity for vector  $g_1, g_2$**

$$\begin{aligned} g_1 &= [g_{1,1}, g_{1,2}, \dots, g_{1,10}] \\ g_2 &= [g_{2,1}, g_{2,2}, \dots, g_{2,10}] \end{aligned}$$

$$similarity(g_1, g_2) = \sum_{i=1}^{10} \frac{1}{\|g_1\| \|g_2\|} (g_{1,i} \cdot g_{2,i})$$

## Intro

---

Identifying gene networks facilitates research on drug mechanisms and transcription factors. Existing studies rather focus on individual genes. We took network approach due to interaction among genes. We used Matrix Factorization to extract properties of gene. It also facilitates analyzing datasets of different platforms or data with outliers.

## Our approach

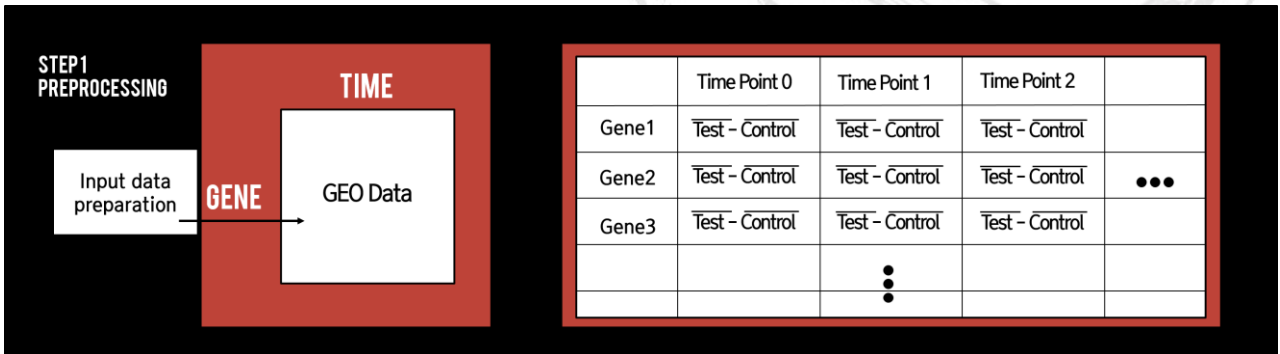
---

1. Apply Matrix Factorization (Biased MF) to dataset
2. Utilize GEO data (GSE 84094, GSE84095)
3. Use data of Non-Small Cell Lung Cancer of human for validation

# Implementation

## STEP 1(PREPROCESSING)

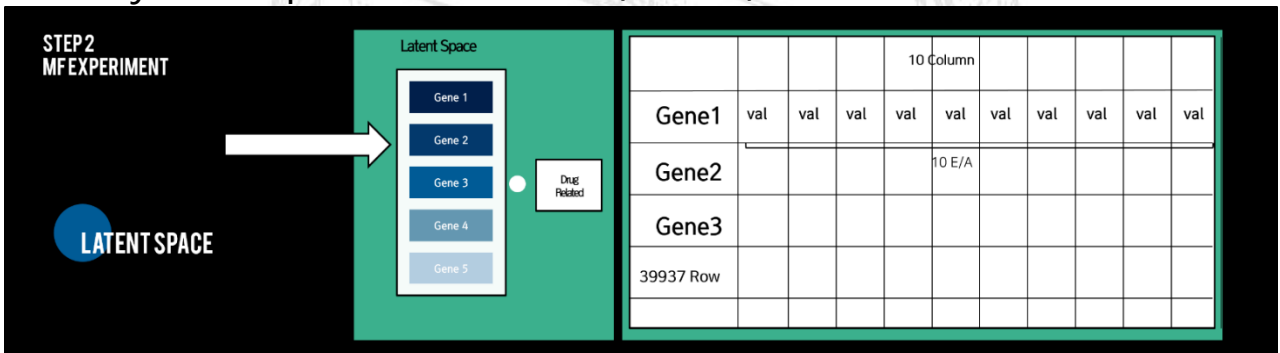
Constructing a matrix using GEO data.



## STEP 2(MF EXPERIMENT)

Obtaining a latent space of 10 columns

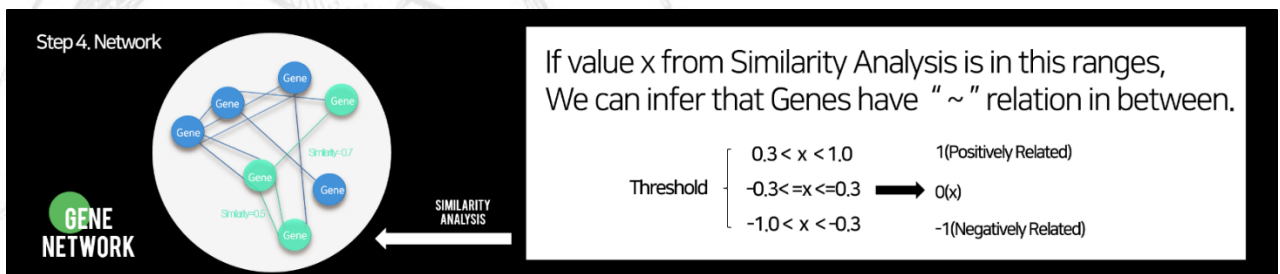
We adjusted open-source code (LibRec) for Matrix Factorization.



## STEP 3(SIMILARITY ANALYSIS)

The similarity between the genes were measured by normalized cosine similarity.

## STEP 4(NETWORK CONSTRUCTION)





We compared the network to existing gene pathway by calculating false negative(type 1 error).

## Ex)

A network diagram illustrating gene-gene interactions. The nodes are labeled "Gene" and are colored blue or green. The edges represent interactions, with some labeled "Similarity=0.7" and "Similarity=0.5".

Compare  
our model with  
gene pathway

The diagram illustrates the signaling pathways involved in the progression of lung cancer, starting from normal epithelial cells and progressing through various stages of malignancy. The pathways shown include:

- Normal cells:** Alveolar and bronchial epithelial cells, Bronchial epithelial cells.
- Atypical adenomatous hyperplasia:** Bronchial dysplasia.
- Primary adenocarcinoma:** Primary squamous cell carcinoma.
- Metastatic adenocarcinoma:** Metastatic squamous cell carcinoma.

Key signaling molecules and pathways depicted include:

- EGFR pathway:** EGFR (E7060, E7061, E7062, E7063, E7064, E7065, E7066, E7067, E7068, E7069, E7070, E7071, E7072, E7073, E7074, E7075, E7076, E7077, E7078, E7079, E7080, E7081, E7082, E7083, E7084, E7085, E7086, E7087, E7088, E7089, E7090, E7091, E7092, E7093, E7094, E7095, E7096, E7097, E7098, E7099, E7100, E7101, E7102, E7103, E7104, E7105, E7106, E7107, E7108, E7109, E7110, E7111, E7112, E7113, E7114, E7115, E7116, E7117, E7118, E7119, E7120, E7121, E7122, E7123, E7124, E7125, E7126, E7127, E7128, E7129, E7130, E7131, E7132, E7133, E7134, E7135, E7136, E7137, E7138, E7139, E7140, E7141, E7142, E7143, E7144, E7145, E7146, E7147, E7148, E7149, E7150, E7151, E7152, E7153, E7154, E7155, E7156, E7157, E7158, E7159, E7160, E7161, E7162, E7163, E7164, E7165, E7166, E7167, E7168, E7169, E7170, E7171, E7172, E7173, E7174, E7175, E7176, E7177, E7178, E7179, E7180, E7181, E7182, E7183, E7184, E7185, E7186, E7187, E7188, E7189, E7190, E7191, E7192, E7193, E7194, E7195, E7196, E7197, E7198, E7199, E7200, E7201, E7202, E7203, E7204, E7205, E7206, E7207, E7208, E7209, E7210, E7211, E7212, E7213, E7214, E7215, E7216, E7217, E7218, E7219, E7220, E7221, E7222, E7223, E7224, E7225, E7226, E7227, E7228, E7229, E7230, E7231, E7232, E7233, E7234, E7235, E7236, E7237, E7238, E7239, E7240, E7241, E7242, E7243, E7244, E7245, E7246, E7247, E7248, E7249, E7250, E7251, E7252, E7253, E7254, E7255, E7256, E7257, E7258, E7259, E7260, E7261, E7262, E7263, E7264, E7265, E7266, E7267, E7268, E7269, E7270, E7271, E7272, E7273, E7274, E7275, E7276, E7277, E7278, E7279, E7280, E7281, E7282, E7283, E7284, E7285, E7286, E7287, E7288, E7289, E7290, E7291, E7292, E7293, E7294, E7295, E7296, E7297, E7298, E7299, E7300, E7301, E7302, E7303, E7304, E7305, E7306, E7307, E7308, E7309, E7310, E7311, E7312, E7313, E7314, E7315, E7316, E7317, E7318, E7319, E7320, E7321, E7322, E7323, E7324, E7325, E7326, E7327, E7328, E7329, E7330, E7331, E7332, E7333, E7334, E7335, E7336, E7337, E7338, E7339, E7340, E7341, E7342, E7343, E7344, E7345, E7346, E7347, E7348, E7349, E7350, E7351, E7352, E7353, E7354, E7355, E7356, E7357, E7358, E7359, E7360, E7361, E7362, E7363, E7364, E7365, E7366, E7367, E7368, E7369, E7370, E7371, E7372, E7373, E7374, E7375, E7376, E7377, E7378, E7379, E7380, E7381, E7382, E7383, E7384, E7385, E7386, E7387, E7388, E7389, E7390, E7391, E7392, E7393, E7394, E7395, E7396, E7397, E7398, E7399, E7400, E7401, E7402, E7403, E7404, E7405, E7406, E7407, E7408, E7409, E7410, E7411, E7412, E7413, E7414, E7415, E7416, E7417, E7418, E7419, E7420, E7421, E7422, E7423, E7424, E7425, E7426, E7427, E7428, E7429, E7430, E7431, E7432, E7433, E7434, E7435, E7436, E7437, E7438, E7439, E7440, E7441, E7442, E7443, E7444, E7445, E7446, E7447, E7448, E7449, E7450, E7451, E7452, E7453, E7454, E7455, E7456, E7457, E7458, E7459, E7460, E7461, E7462, E7463, E7464, E7465, E7466, E7467, E7468, E7469, E7470, E7471, E7472, E7473, E7474, E7475, E7476, E7477, E7478, E7479, E7480, E7481, E7482, E7483, E7484, E7485, E7486, E7487, E7488, E7489, E7490, E7491, E7492, E7493, E7494, E7495, E7496, E7497, E7498, E7499, E7500, E7501, E7502, E7503, E7504, E7505, E7506, E7507, E7508, E7509, E7510, E7511, E7512, E7513, E7514, E7515, E7516, E7517, E7518, E7519, E7520, E7521, E7522, E7523, E7524, E7525, E7526, E7527, E7528, E7529, E7530, E7531, E7532, E7533, E7534, E7535, E7536, E7537, E7538, E7539, E7540, E7541, E7542, E7543, E7544, E7545, E7546, E7547, E7548, E7549, E7550, E7551, E7552, E7553, E7554, E7555, E7556, E7557, E7558, E7559, E7560, E7561, E7562, E7563, E7564, E7565, E7566, E7567, E7568, E7569, E7570, E7571, E7572, E7573, E7574, E7575, E7576, E7577, E7578, E7579, E7580, E7581, E7582, E7583, E7584, E7585, E7586, E7587, E7588, E7589, E7590, E7591, E7592, E7593, E7594, E7595, E7596, E7597, E7598, E7599, E7600, E7601, E7602, E7603, E7604, E7605, E7606, E7607, E7608, E7609, E7610, E7611, E7612, E7613, E7614, E7615, E7616, E7617, E7618, E7619, E7620, E7621, E7622, E7623, E7624, E7625, E7626, E7627, E7628, E7629, E7630, E7631, E7632, E7633, E7634, E7635, E7636, E7637, E7638, E7639, E7640, E7641, E7642, E7643, E7644, E7645, E7646, E7647, E7648, E7649, E7650, E7651, E7652, E7653, E7654, E7655, E7656, E7657, E7658, E7659, E7660, E7661, E7662, E7663, E7664, E7665, E7666, E7667, E7668, E7669, E7670, E7671, E7672, E7673, E7674, E7675, E7676, E7677, E7678, E7679, E7680, E7681, E7682, E

As a result of validation test, Type 1 Error occurred. Comparing to other models that predict the pathways with moderate precision, our model could predict only about 34% of existing pathways, which requires improvement on precision. Adding extra columns in original matrix is being considered to improve the precision of the MF method.