

Moran I Coefficient Fluctuation Analysis for FYN Variations in Adherens Junction

Jeff Cromwell, PhD

Mathematical Learning Space Research Portfolio Spring Semester 2021: (DRAFT) Preprint 78

1 Abstract

Signaling pathways for FYN a proto-oncogene at 6q21 is a Src family tyrosine kinase with orthology "tyrosine-protein kinase Fyn [EC:2.7.10.2]" with pathway (a) Sphingolipid signaling pathway (b) Phospholipase D signaling pathway (c) Axon guidance (d) Osteoclast differentiation (e) Focal adhesion (f) Adherens junction" (g) Platelet activation cell mediated cytotoxicity (h) T cell receptor signaling pathway (i) Fc epsilon RI signaling pathway (j) Cholinergic synapse (k) Prion disease (l) Pathogenic Escherichia coli infection (l) Viral myocarditis with networks of (a) Cytoskeletal regulation (viruses and bacteria), (b) PI3K signaling and Prion disease with network elements of Escherichia EspJ/Tir to Actin signaling pathway PRNP-PI3K-NOX2 signaling pathway and Scrapie conformation PrPSc to PRNP-PI3K-NOX2 signaling pathway.

Variations in the structure of FYN with 30 different structures is examined here with Hasegawa, Kishino and Yano (HKY85) model and properties examined with the Protein Stability Index Binding Potential ALiphatic f.1 CpH5 CpH7 and CpH9 and ten Kidera factors. The first four factors are essentially pure physical properties; the remaining six factors are superpositions of several physical properties as well as Moran I Coefficient Fluctuation Analysis.

Keywords : tyrosine-protein kinase proto-oncogene Moran I Coefficient 6IPY 4U1P 3UA7 6IPZ 4EIK 3H0H 3UA6 1SHF 4U17 4ZNX 3H0I 1FYN 1M27 1EFN 4D8D 1G83 3H0F 2DQ7 1AVZ 5ZAU 2MRJ 2MRK 2MQI 1ZBJ 1A0N 1AZG 1AOT 1AOU 1NYF 1NYG

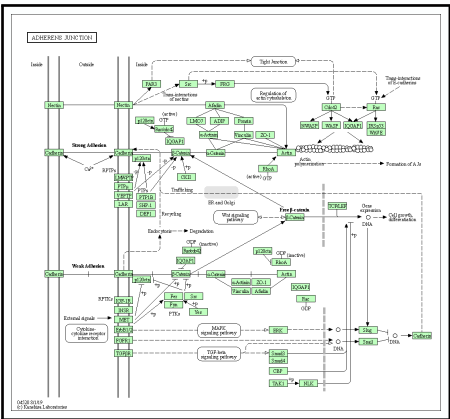
2 Introduction

*The Mathematical Learning Space Research Portfolio
Email address: <http://mathlearningspace.weebly.com/> (Jeff Cromwell, PhD)

FYN Variations are involved in Environmental Information Processing such as Signal transduction with Phospholipase D signaling pathway and Sphingolipid signaling pathway with Cellular Processes Cellular community - eukaryotes as Focal adhesion and Adherens junction. For Organismal Systems and the Immune system (a) Platelet activation (b) cell mediated cytotoxicity (c) T cell receptor signaling pathway (d) Fc epsilon RI signaling pathway with 09156 Nervous system Cholinergic synapse an Development and regeneration with Axon guidance and Osteoclast differentiation. [401]

Human Diseases such as Infectious disease with bacterial like Pathogenic Escherichia coli infection and Neurodegenerative disease with Prion disease, Cardiovascular disease with Viral myocarditis are involved with FYN. Metabolism Protein kinases, genetic information processing Membrane trafficking. Enzymes with 2. Transferases 2.7 Transferring phosphorus-containing groups 2.7.10 Protein-tyrosine kinases 2.7.10.2 non-specific protein-tyrosine kinase are also connected to FYN as well as Protein kinases such as Non-receptor tyrosine kinases SRC family. Membrane trafficking with Exocytosis and Calcium ion-dependent exocytosis Kinases and associated proteins and Endocytosis with Lipid raft mediated endocytosis Flotillin-dependent endocytosis [401]

Figure 1 has the Cell-cell adherens junctions (AJs) is the most common type of intercellular adhesions and important for maintaining tissue architecture and cell polarity and can limit cell movement and proliferation. Changes in the phosphorylation state of beta-catenin affect (a) cell-cell adhesion, (b) cell migration and the (c) level of signaling beta-catenin. Wnt signaling acts as a positive regulator of beta-catenin by inhibiting beta-catenin degradation and stabilizes beta-catenin and causes its accumulation.[401]



$$P(t) = \begin{pmatrix} p_{AA}(t) & p_{AG}(t) & p_{AC}(t) & p_{AT}(t) \\ p_{GA}(t) & p_{GG}(t) & p_{GC}(t) & p_{GT}(t) \\ p_{CA}(t) & p_{CG}(t) & p_{CC}(t) & p_{CT}(t) \\ p_{TA}(t) & p_{TG}(t) & p_{TC}(t) & p_{TT}(t) \end{pmatrix} \quad (2)$$

where the top-left and bottom-right 2x2 blocks correspond to "transition probabilities" and the top-right and bottom-left 2x2 blocks corresponds to "transversion probabilities". [1000] Table 1 has the maximum likelihood fit of sequence 1 and estimate of the eigenvalues from the markov chain transition matrix.

| | 1 | 2 |
|----|---------------------------------------|---------------------------------------|
| 1 | 1+0i | 0.612672547758859+0i |
| 2 | 0.548598035811199+0i | -0.253217867138598+0.43157696760881i |
| 3 | -0.253217867138598-0.43157696760881i | 0.160762332831767+0.457336167791463i |
| 4 | 0.160762332831767-0.457336167791463i | 0.383223041837455+0i |
| 5 | -0.330813409838521+0i | 0.21505677396784+0.243678398698828i |
| 6 | 0.21505677396784-0.243678398698828i | -0.299147890266987+0.125575933714316i |
| 7 | -0.299147890266987-0.125575933714316i | -0.131874050947548+0.195229080875112i |
| 8 | -0.131874050947548-0.195229080875112i | 0.091335297535122+0.201048572422311i |
| 9 | 0.091335297535122-0.201048572422311i | -0.0652236932464657+0i |
| 10 | -1.5288188771906e-16+0i | 1+0i |

4 Results

Table 2 has protein, Stability Index, Binding Potential, ALiphatic, f.1, CpH5, CpH7 andCpH9. [1001]

| | Protein | Stability Index | Binding Potential | ALiphatic | f.1 | CpH5 | CpH7 | CpH9 |
|----|---------|-----------------|-------------------|-----------|-----|--------|--------|---------|
| 1 | 6IPY | 37.4 | 1.43 | 81.5 | 1 | -4.90 | -8.29 | -10.032 |
| 2 | 4U1P | 31.1 | 2.23 | 71.5 | 1 | 8.30 | 2.33 | -2.676 |
| 3 | 3UA7 | 41.0 | 1.66 | 69.2 | 1 | -17.00 | -24.03 | -27.008 |
| 4 | 6IPZ | 34.9 | 1.51 | 68.1 | 1 | -5.95 | -7.78 | -9.038 |
| 5 | 4EIK | 55.0 | 1.62 | 80.6 | 1 | -2.03 | -3.78 | -5.039 |
| 6 | 3HOH | 34.9 | 1.33 | 74.3 | 1 | -5.95 | -7.78 | -9.038 |
| 7 | 3UA6 | 39.0 | 1.73 | 70.9 | 1 | -9.02 | -12.53 | -14.361 |
| 8 | 1SHF | 37.4 | 1.76 | 72.7 | 1 | -9.95 | -13.53 | -15.361 |
| 9 | 4U17 | 28.3 | 2.34 | 69.4 | 1 | 30.97 | 17.84 | 5.781 |
| 10 | 4ZNX | 53.2 | 1.91 | 68.6 | 1 | -8.22 | -15.03 | -18.009 |
| 11 | 3HOI | 37.4 | 1.43 | 79.3 | 1 | -11.95 | -15.53 | -17.360 |
| 12 | 1FYN | 50.5 | 1.38 | 65.0 | 1 | -4.95 | -6.78 | -8.112 |
| 13 | 1M27 | 28.4 | 1.38 | 86.9 | 1 | 2.92 | -2.12 | -6.426 |
| 14 | 1EFN | 41.4 | 1.57 | 79.3 | 1 | 7.62 | -6.70 | -14.003 |
| 15 | 4D8D | 38.2 | 1.43 | 78.0 | 1 | 5.62 | -8.70 | -16.033 |
| 16 | 1G83 | 34.2 | 2.22 | 72.1 | 1 | 4.32 | -6.07 | -10.449 |
| 17 | 3HOF | 34.8 | 1.47 | 72.7 | 1 | -5.95 | -7.78 | -9.038 |
| 18 | 2DQ7 | 40.4 | 1.45 | 89.7 | 1 | 5.37 | -1.16 | -6.852 |
| 19 | 1AVZ | 39.8 | 1.51 | 79.1 | 1 | 9.68 | -2.94 | -9.679 |
| 20 | 5ZAU | 33.3 | 1.74 | 71.7 | 1 | 3.75 | 1.23 | -0.462 |
| 21 | 2MRJ | 29.2 | 2.33 | 71.2 | 1 | 10.92 | 6.85 | 2.449 |
| 22 | 2MRK | 31.3 | 2.35 | 68.3 | 1 | 9.14 | 4.85 | 0.375 |
| 23 | 2MQI | 29.2 | 2.33 | 71.2 | 1 | 10.92 | 6.85 | 2.449 |
| 24 | 1ZBJ | 37.4 | 1.76 | 72.7 | 1 | -4.95 | -6.78 | -8.038 |
| 25 | 1AON | 49.5 | 1.58 | 70.4 | 1 | -2.03 | -3.78 | -5.054 |
| 26 | 1AZG | 49.5 | 1.58 | 70.4 | 1 | -2.03 | -3.78 | -5.054 |
| 27 | 1AOT | 36.9 | 2.34 | 75.0 | 1 | 6.48 | 1.95 | -0.419 |
| 28 | 1AOU | 36.9 | 2.34 | 75.0 | 1 | 6.48 | 1.95 | -0.419 |
| 29 | 1NYF | 40.4 | 1.64 | 74.0 | 1 | -4.03 | -5.78 | -7.038 |
| 30 | 1NYG | 40.4 | 1.64 | 74.0 | 1 | -4.03 | -5.78 | -7.038 |

Table 1: A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable. A protein have high binding potential if the index value is higher than 2.48. The relative volume occupied by aliphatic side chains (Alanine, Valine, Isoleucine, and Leucine) is a positive factor for the increase of thermostability of globular proteins. net charge of a protein sequence based on the Henderson-Hasselbalch equation for pH levels of 5 7 and 9. f.1 is the crosscovariance based on one of the sequence S at lag 1 is given by f.1 with property1 = Hydrophobicity based on the KyteDoolittle Scale and property2 =Hydrophobicity based on the Eisenberg scale.

4.1. Moran I Coefficient Fluctuation Analysis

The matrix weight is used as "neighbourhood" weights, and Moran's I coefficient is computed using the formula: [1007][1008]

$$I = n/S0 * (\sum_{i=1..n} \sum_{j=1..n} w_{ij}(y_i - y_m)(y_j - y_m)) / (\sum_{i=1..n} (y_i - y_m)^2) \quad (3)$$

with y_i = observations, w_{ij} = distance weight n = number of observations $S0 = \sum_{i=1..n} \sum_{j=1..n} w_{ij}$. The null hypothesis of no phylogenetic correlation is tested assuming normality of I under this null hypothesis. If the observed value of I is significantly greater than the expected value, then the values of x are positively autocorrelated, whereas if $I_{observed} < I_{expected}$, this will indicate negative autocorrelation. Here observed the computed Moran's I . expected the expected value of I under the null hypothesis. sd the standard deviation of I under the null hypothesis. p.value the P-value of the test of the null hypothesis against the alternative hypothesis specified in alternative is shown in Table 3. [1]

| | FYN | Observed | Expected | SD | PValue |
|----|------|----------|----------|---------|----------|
| 1 | 6IPY | -0.00158 | -0.0152 | 0.0104 | 0.192 |
| 2 | 4U1P | 0.0333 | -0.0087 | 0.00926 | 5.6e-06 |
| 3 | 3UA7 | 0.0366 | -0.00383 | 0.00411 | 0 |
| 4 | 6IPZ | 0.0235 | -0.0161 | 0.0131 | 0.00256 |
| 5 | 4EIK | 0.0722 | -0.0147 | 0.0191 | 5.22e-06 |
| 6 | 3HOH | 0.0175 | -0.0161 | 0.00739 | 5.24e-06 |
| 7 | 3UA6 | 0.0163 | -0.00862 | 0.0085 | 0.00342 |
| 8 | 1SHF | 0.0133 | -0.00855 | 0.0085 | 0.01 |
| 9 | 4U17 | 0.0405 | -0.00326 | 0.00375 | 0 |
| 10 | 4ZNX | 0.0281 | -0.00377 | 0.00534 | 2.29e-09 |
| 11 | 3HOI | 0.0187 | -0.00855 | 0.00948 | 0.00411 |
| 12 | 1FYN | 0.0217 | -0.0141 | 0.0145 | 0.0135 |
| 13 | 1M27 | 0.0186 | -0.00571 | 0.00459 | 1.23e-07 |
| 14 | 1EFN | 0.0301 | -0.00312 | 0.00321 | 0 |
| 15 | 4D8D | 0.0273 | -0.00308 | 0.00322 | 0 |
| 16 | 1G83 | 0.0721 | -0.00312 | 0.00549 | 0 |
| 17 | 3HOF | 0.0302 | -0.0172 | 0.0147 | 0.00123 |
| 18 | 2DQ7 | 0.021 | -0.00382 | 0.00307 | 6.66e-16 |
| 19 | 1AVZ | 0.0327 | -0.00386 | 0.00424 | 0 |
| 20 | 5ZAU | 0.0243 | -0.00833 | 0.00771 | 2.31e-05 |
| 21 | 2MRJ | 0.0376 | -0.0101 | 0.011 | 1.54e-05 |
| 22 | 2MRK | 0.0253 | -0.00917 | 0.00842 | 4.27e-05 |
| 23 | 2MQI | 0.0496 | -0.0101 | 0.00938 | 1.94e-10 |
| 24 | 1ZBJ | 0.0218 | -0.0172 | 0.0266 | 0.142 |
| 25 | 1AON | 0.0418 | -0.0141 | 0.0128 | 1.25e-05 |
| 26 | 1AZG | 0.0651 | -0.0141 | 0.0127 | 4.14e-10 |
| 27 | 1AOT | 0.0281 | -0.00862 | 0.00478 | 1.62e-14 |
| 28 | 1AOU | 0.0232 | -0.00862 | 0.00484 | 5.08e-11 |
| 29 | 1NYF | 0.0266 | -0.0175 | 0.0178 | 0.0129 |
| 30 | 1NYG | 0.0142 | -0.0175 | 0.0294 | 0.28 |

Various fixed width running window sizes can be used for summaries such as: (1) runsum, (2) runmean, (3) runmed, (4) runwtsun, (5) runq functions calculate the sum, mean, median, weighted sum, and order statistic. Figure 2 has the moving window first moment with 4ZNX, 5ZAU, and 1NYG. In the figure, the differences among this group is demonstrated with variation in the length 3 8 15 24 and 50 for the window. [1000]

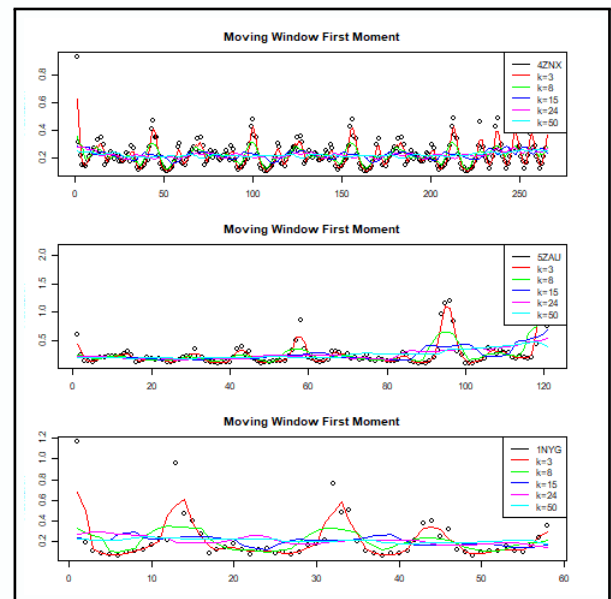


Figure 2: Moving Window First Moment for (a) 4ZNX (b) 5ZAU (c) 1NYG

Figure 3 has the Pearson Types for FYN Variations.[1000]

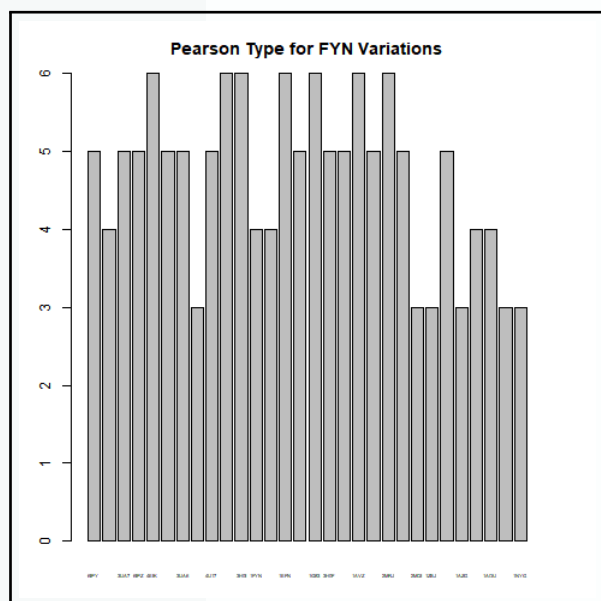


Figure 3: Pearson Types for FYN Variations

Figure 4 has the Kurtosis Values for the FYN Variations.[1000]

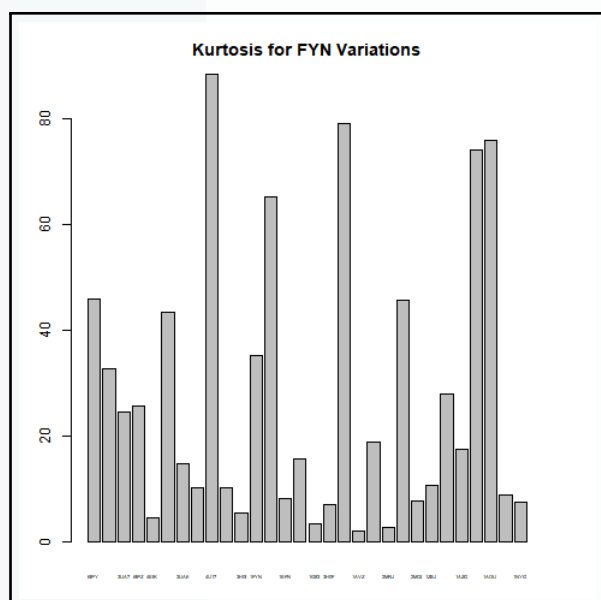


Figure 4: Kurtosis Values for the FYN Variations

5 Conclusions

Variations in the structure of FYN with 30 different structures is examined here with HKY85 model by Hasegawa, Kishino and Yano 1985 and properties examined with the Protein Stability Index Binding Potential ALiphatic f.1 CpH5 CpH7 and CpH9 and ten Kidera factors. The first four

factors are essentially pure physical properties; the remaining six factors are superpositions of several physical properties as well as Moran I Coefficient Fluctuation Analysis. In the moving window first moment with 4ZNX, 5ZAU, and 1NYG the variation in the length 3 8 15 24 and 50 for the window.

6 References

- [1] Wikipedia contributors. "Moran's I." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 11 Aug. 2021. Web. 15 Aug. 2021.
- [10] Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205 (2014).
- [101] Wikipedia contributors. "Graph theory." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 24 Mar. 2021. Web. 24 Mar. 2021.
- [102] Wikipedia contributors. "Spectral graph theory." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 19 Mar. 2021. Web. 24 Mar. 2021.
- [400] Kanehisa, Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K.; KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361 (2017).
- [401] Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M.; KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462 (2016).
- [402] Kanehisa, M. and Goto, S.; KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 28, 27–30 (2000).
- [420] GeneCards Version 3: the human gene integrator
- [430] DrugBank 5.0: a major update to the DrugBank database for 2018.
- [440] COSMIC: the Catalogue Of Somatic Mutations In Cancer.
- [450] Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders
- [460] The ClinicalTrials.gov Results Database — Update and Key Issues
- [470] PubChem Substance and Compound databases
- [480] The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible
- [490] MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search
- [803] Cromwell, J. "Mathematical Learning Space Research Portfolio" Mathematical Learning Space Research Portfolio, <http://mathlearningspace.weebly.com/> 8 3 2021. Web. 3 Aug. 2021.
- [1000] R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [1001] Osorio, D., Rondon-Villarreal, P. and Torres, R. Peptides: A package for data mining of antimicrobial peptides. *The R Journal.* 7(1), 4–14 (2015).
- [1002] Spedicato Giorgio Alfredo Discrete Time Markov Chains with R. *The R Journal* 7 (2017).
- [1003] Ingo Feinerer and Kurt Hornik (2017). tm: Text Mining Package. R package version 0.7-1. <https://CRAN.R-project.org/package=tm>
- [1004] Ingo Feinerer, Kurt Hornik, and David Meyer (2008). Text Mining Infrastructure in R. *Journal of Statistical Software* 25(5): 1–54. URL: <http://www.jstatsoft.org/v25/i05/>.
- [1005] Sarkar, Deepayan (2008) Lattice: Multivariate Data Visualization with R. Springer, New York. ISBN 978-0-387-75968-5

- [1006] Luna A et al. rcellminer: Exploring Molecular Profiles and Drug Response of the NCI-60 Cell Lines in R. Bioinformatics. 2015 Dec. <http://www.ncbi.nlm.nih.gov/pubmed/26635141>
- [1007] Roger Bivand, Gianfranco Piras (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. Journal of Statistical Software, 63(18), 1-36. URL <http://www.jstatsoft.org/v63/i18/>.
- [1008] Bivand, R. S., Hauke, J., and Kossowski, T. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. Geographical Analysis, 45(2), 150-179