BIBLIOGRAPHY

Akey, J.M., et al. "Population history and natural selection shape patterns of genetic variation in 132 genes." *PLoS Biology.* 2 (2004) : 286.

Alcazar, A., et al. "Differentiation of green, white, black, oolong, and Pu-erh teas according to their free amino acids content." *Journal of Agricultural and Food Chemistry.* 55 (2007) : 5960–5965.

Amendolia, S.R., et al. "A real-time classification system of thalassemic pathologies based on artificial neural networks." *Medical Decision Making.* 22 (2002) : 18–26.

Amendolia, S.R., et al. "A comparative study of k-nearest neighbour, support vector machine and multi-layer perceptron for thalassemia screening." *Chemometrics and Intelligent Laboratory Systems.* 69 (2003) : 13–20.

Bamshad, M. and Wooding, S.P. "Signatures of natural selection in the human genome." *Nature Reviews Genetics.* 4 (2003) : 99–111.

Barosi, G., et al. "Classification of anemia on the basis of ferrokinetic parameters." *British Journal of Haematology.* 61 (1985) : 357–370.

Baye, T.M., el al. "Database mining for selection of SNP markers useful in admixture mapping." *BioData Mining.* 2 (2009) : 1.

Beaumont, M.A. and Balding, D.J. "Identifying adaptive genetic divergence among populations from genome scans." *Molecular Ecology.* 13 (2004) : 969–980.

Birndorf, N.I., et al. "An expert system to diagnose anemia and report results directly on hematology forms." *Computers and Biomedical Research.* 29 (1996) : 16–26.

Breiman, L., et al. *Classification and Regression Trees.* Belmont, CA : Wadsworth International Group, 1984.

Breiman, L. "Random forests." *Machine Learning*. 45 (2001) : 5–32.

Bryk, J., et al. "Positive selection in East Asians for an *EDAR* allele that enhances NF-κB activation." *PLoS One*. 3 (2008) : 2209.

Budowle, B. and van Daal, A. "Forensically relevant SNP classes." *BioTechniques*. 44 (2008) : 603–610.

Cheng, F., et al. "SNP@Evolution: a hierarchical database of positive selection on the human genome." *BMC Evolutionary Biology*. 9 (2009) : 221.

Chuang, C.-C., Wen, W.-C. and Sheu, S.-J. "Classification of Aurantii Fructus samples by multivariate analysis." *Journal of Separation Science*. 30 (2007) : 1827–1832.

Clarke, G.M. and Higgins, T.N., "Laboratory investigation of hemoglobinopathies and thalassemias: Review and update." *Clinical Chemistry*. 46 (2000) : 1284–1290.

Colah, R.B., et al. "HPLC studies in hemoglobinopathies." *Indian Journal of Pediatrics*. 74 (2007) : 657–662.

Demir, A., et al. "Most reliable indices in differentiation between thalassemia trait and iron deficiency anemia." *Pediatrics International*. 44 (2002) : 612–616.

Donald, D., et al. "Adaptive wavelet modelling of a nested 3 factor experimental design in NIR chemometrics." *Chemometrics and Intelligent Laboratory Systems*. 82 (2006a) : 122–129.

Donald, D., et al. "Bagged super wavelets reduction for boosted prostate cancer classification of seldi-tof mass spectral serum profiles." *Chemometrics and Intelligent Laboratory Systems*. 82 (2006b) : 2–7.

Engle, R.L., et al. "HEME: A computer aid to diagnosis of hematologic disease." *Bulletin of the New York Academy of Medicine*. 52 (1976) : 584–600.

Enoch, M.A., et al. "Using ancestry-informative markers to define populations and detect population stratification." *Journal of Psychopharmacology*. 20 (2006) : 19–26.

Falush, D., Stephens, M. and Pritchard, J.K. "Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies." *Genetics*. 164 (2003) : 1567–1587.

_____. "Inference of population structure using multilocus genotype data: dominant markers and null alleles." *Molecular Ecology Notes*. 7 (2007) : 574–578.

Fayyad, U.M. and Irani, K.B. "Multi-interval discretization of continuous-valued attributes for classification learning." Edited by Bajcsy, R. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence, Chambéry, France*. San Mateo, CA : Morgan Kaufmann, 1993 : 1022–1027.

Fucharoen, S. and Winichagoon, P. "Hemoglobinopathies in Southeast Asia: Molecular biology and clinical medicine." *Hemoglobin*. 21 (1997) : 299–319.

Fujimoto, A., et al. "A scan for genetic determinants of human hair morphology: *EDAR* is associated with Asian hair thickness." *Human Molecular Genetics*. 17 (2008) : 835–843.

Gao, X. and Starmer, J. "Human population structure detection via multilocus genotype clustering." *BMC Genetics*. 8 (2007) : 34.

Granitto, P.M., et al. "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products." *Chemometrics and Intelligent Laboratory Systems*. 83 (2006) 83–90.

Hall, M.A. and Holmes, G. "Benchmarking attribute selection techniques for discrete class data mining." *IEEE Transactions on Knowledge and Data Engineering*. 15 (2003) : 1437–1447.

Hancock, T., et al. "A performance comparison of modern statistical techniques for molecular descriptor selection and retention prediction in chromatographic QSRR studies." *Chemometrics and Intelligent Laboratory Systems*. 76 (2005) : 185–196.

Hartwell, L.H., et al. *Genetics: From Genes to Genomes*. 3rd ed. New York : McGraw-Hill, 2008.

Hastie, T. and Tibshirani, R. "Classification by pairwise coupling." *Annals of Statistics*. 26 (1998) : 451–471.

Hinds, D.A., et al. "Whole-genome patterns of common DNA variation in three human populations." *Science*. 307 (2005) : 1072–1079.

Jimenez, C.V., Minchinela, J. and Ros, J. "New indices from the H*2 analyser improve differentiation between heterozygous beta or delta beta thalassaemia and iron-deficency anaemia." *Clinical and Laboratory Haematology*. 17 (1995) : 151–155.

Jobling, M.A. "In the name of the father: surnames and genetics." *Trends in Genetics*. 17 (2001) : 353–357.

Jobling, M.A. and Tyler-Smith, C. "The Human Y chromosome: An evolutionary marker comes of age." *Nature Reviews Genetics*. 4 (2003) : 598–612.

Joutovsky, A., Hadzi-Nesic, J. and Nardi, M.A. "HPLC retention time as a diagnostic tool for hemoglobin variants and hemoglobinopathies: A study of 60,000 samples in a clinical diagnostic laboratory." *Clinical Chemistry*. 50 (2004) : 1736–1747.

Karafet, T.M., et al. "New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree." *Genome Research*. 18 (2008) : 830–838.

Kher, A., et al. "Forensic classification of ballpoint pen inks using high performance liquid chromatography and infrared spectroscopy with principal components analysis and linear discriminant analysis." *Vibrational Spectroscopy*. 40 (2006) : 270–277.

Kelley, J.L., et al. "Genomic signatures of positive selection in humans and the limits of outlier approaches." *Genome Research*. 16 (2006) : 980–989.

Kohavi, R. "A study of cross-validation and bootstrap for accuracy estimation and model selection." In *Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montréal, Québec, Canada*. San Mateo, CA : Morgan

Kaufmann, 1995 : 1137–1143.

Kohavi, R. and John, G.H. "Wrappers for feature subset selection." *Artificial Intelligence.* 97 (1997) : 273–324.

Kosoy, R., et al. "Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America." *Human Mutation.* 30 (2009) : 69–78.

Kutlar, F. "Diagnostic approach to hemoglobinopathies." *Hemoglobin.* 31 (2007) : 243–250.

Lanzola, G., et al. "NEOANEMIA: A knowledge-based system emulating diagnostic reasoning." *Computers and Biomedical Research.* 23 (1990) : 560–582.

Lao, O., et al. "Proportioning whole-genome single-nucleotide-polymorphism diversity for the identification of geographic population structure and genetic ancestry." *American Journal of Human Genetics.* 78 (2006) : 680–690.

Lewin, R. "Africa: Cradle of modern humans." *Science.* 237 (1987) : 1292–1295.

López Herráez, D., et al. "Genetic variation and recent positive selection in worldwide human populations: Evidence from nearly 1 million SNPs." *PLoS One.* 4 (2009) : 7888.

Lund, P.R. and Barnes, R.D. "Automated classification of anaemia using image analysis." *The Lancet.* 300 (1972) : 463–464.

Mellars, P. "Why did modern human populations disperse from Africa *ca.* 60,000 years ago? A new model." *Proceedings of the National Academy of Sciences of the United States of America.* 103 (2006) : 9381–9386.

Metspalu, M., et al. "Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans." *BMC Genetics.* 5 (2004) : 26.

Mitchell, T.M. *Machine Learning.* Singapore : McGraw-Hill, 1997.

Moore, J.H., et al. "A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human

disease susceptibility." *Journal of Theoretical Biology.* 241 (2006) : 252–261.

Mou, C., et al. "Enhanced ectodysplasin-A receptor (*EDAR*) signaling alters multiple fiber characteristics to produce the East Asian hair form." *Human Mutation.* 29 (2008) : 1405–1411.

Mulholland, M., et al. "A comparison of classification in artificial intelligence, induction versus a self-organising neural networks." *Chemometrics and Intelligent Laboratory Systems.* 30 (1995a) : 117–128.

Mulholland, M., et al. "Application of the C4.5 classifier to building an expert system for ion chromatography." *Chemometrics and Intelligent Laboratory Systems.* 27 (1995b) : 95–104.

Myles, S., et al. "Identification and analysis of genomic regions with large between-population differentiation in humans." *Annals of Human Genetics.* 72 (2008) : 99–110.

Ntaios, G., et al. "Discrimination indices as screening tests for β-thalassemic trait." *Annals of Hematology.* 86 (2007) : 487–491.

Nunkesser, R., et al. "Detecting high-order interactions of single nucleotide polymorphisms using genetic programming." *Bioinformatics.* 23 (2007) : 3280–3288.

Old, J.M. "Screening and genetic diagnosis of haemoglobin disorders." *Blood Reviews.* 17 (2003) : 43–53.

Oleksyk, T.K., et al. "Identifying selected regions from heterozygosity and divergence using a light-coverage genomic dataset from two human populations." *PLoS One.* 3 (2008) : 1712.

Olson, S. "Seeking the signs of selection." *Science.* 298 (2002) : 1324–1325.

Ou, C.-N. and Rognerud, C.L. "Diagnosis of hemoglobinopathies: Electrophoresis vs. HPLC." *Clinica Chimica Acta.* 313 (2001) : 187–194.

Park, J., et al. "SNP@Ethnos: A database of ethnically variant single-nucleotide polymorphisms." *Nucleic Acids Research.* 35 (2007) : 711–715.

Paschou, P., et al. "PCA-correlated SNPs for structure identification in worldwide human populations." *PLoS Genetics.* 3 (2007) : 60.

Phillips, C., et al. "Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs." *Forensic Science International: Genetics.* 1 : (2007) 273–280.

Pickrell, J.K., et al. "Signals of recent positive selection in a worldwide sample of human populations." *Genome Research.* 19 (2009) : 826–837.

Pritchard, J.K., Stephens, M. and Donnelly, P. "Inference of population structure using multilocus genotype data." *Genetics.* 155 (2000) : 945–959.

Press, W.H., et al. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge : Cambridge University Press, 1998.

Quaglini, S., et al. "ANEMIA: An expert consultation system." *Computers and Biomedical Research.* 19 (1986) : 13–27.

_____. "A performance evaluation of the expert system ANEMIA." *Computers and Biomedical Research.* 21 (1988) : 307–323.

Quinlan, J.R. *C4.5: Programs for Machine Learning.* San Mateo, CA : Morgan Kaufmann, 1993.

Quintana-Murci, L., et al. "Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through Eastern Africa." *Nature Genetics.* 23 (1999) : 437–441.

Rosenberg, N.A. "Algorithms for selecting informative marker panels for population assignment." *Journal of Computational Biology.* 12 (2005) : 1183–1201.

Rosenberg, N.A., et al. "Informativeness of genetic markers for inference of ancestry." *American Journal of Human Genetics.* 73 (2003) : 1402–1422.

Rumelhart, D.E. and McClelland (Eds.), J.L. "Parallel distributed processing: Explorations in the microstructure of cognition." *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. I.* Cambridge, MA : MIT Press, 1986.

Sabeti, P.C., et al. "Detecting recent positive selection in the human genome from

haplotype structure." *Nature.* 419 (2002) : 832–837.

Sabeti, P.C., et al. "Genome-wide detection and characterization of positive selection in human populations." *Nature.* 449 (2007) : 913–918.

Saeys, Y., Inza, I. and Larrañaga, P. "A review of feature selection techniques in bioinformatics." *Bioinformatics.* 23 (2007) : 2507–2517.

Salas, A., et al. "The making of the African mtDNA landscape." *American Journal of Human Genetics.* 71 (2002) : 1082–1111.

Seldin, M.F. "Admixture mapping as a tool in gene discovery." *Current Opinion in Genetics and Development.* 17 (2007) : 177–181.

The International HapMap Consortium. "The international HapMap project." *Nature.* 426 (2003) : 789–796.

_____. "A haplotype map of the human genome." *Nature.* 437 (2005) : 1299–1320.

_____. "A second generation human haplotype map of over 3.1 million SNPs." *Nature.* 449 (2007) : 851–861.

Tian, C., Gregersen, P.K. and Seldin, M.F. "Accounting for ancestry: Population substructure and genome-wide association studies." *Human Molecular Genetics.* 17 (2008) : 143–150.

Vallender, E.J. and Lahn, B.T. "Positive selection on the human genome." *Human Molecular Genetics.* 13 (2004) : 245–254.

Voight, B.F., et al. "A map of recent positive selection in the human genome." *PLoS Biology.* 4 (2006) : 72.

Weatherall, D.J. and Clegg, J.B. *The Thalassemia Syndromes.* 4th ed. Malden, MA : Blackwell Science, 2001.

Witten, I.H. and Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. San Francisco, CA : Morgan Kaufmann, 2005.

Wongseree, W., et al. "Thalassaemia classification by neural networks and genetic programming." *Information Sciences.* 177 (2007) : 771–786.

Wongseree, W., et al. "Detecting purely epistatic multi-locus interactions by an

omnibus permutation test on ensembles of two-locus analyses." *BMC Bioinformatics*. 10 (2009) : 294.

Wright, S. "The genetical structure of populations." *Annals of Eugenics*. 15 (1995) : 323–354.

Zhao, R.H., et al. "Application of an artificial neural network in chromatography–retention behavior prediction and pattern recognition." *Chemometrics and Intelligent Laboratory Systems*. 45 (1999) : 163–170.

Zhou, N. and Wang, L. "Effective selection of informative SNPs and classification on the HapMap genotype data." *BMC Bioinformatics*. 8 (2007) : 484.

APPENDIX A

The published research results.

The research results have been published as follows.

Piroonratana, T., Wongseree, W., Assawamakin, A., Paulkhaolarn, N., Kanjanakorn, C., Sirikong, M., Thongnoppakhun, W., Limwongse, C. and Chaiyaratana, N."Classification of haemoglobin typing chromatograms by neural networks and decision trees for thalassaemia screening."*Chemometrics and Intelligent Laboratory Systems.* 99 (2009) : 101–110.

Piroonratana, T., Wongseree, W., Usavanarong, T., Assawamakin, A., Limwongse, C. and Chaiyaratana, N."Identification of Ancestry Informative Markers from Chromosome-Wide Single Nucleotide Polymorphisms Using Symmetrical Uncertainty Ranking."In *Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey,*2010 : 2448–2451

BIOGRAPHY

Name : Mr.Theera Piroonratana

Thesis Title : An Information-Theoretic Approach to Pattern Recognition and Its

Application in Life Science

Major : Electrical Engineering

Biography

Personal Information: I was born on February 24, 1981. I currently live in Bangkruai, Nonthaburi, Thailand.

Education: I earned a bachelor degree and a master degree in production engineering from KMITNB.