Dr. Satarupa Banerjee
B.Pharm, M.Tech (Biotechnology), PhD
Institute Postdoctoral Fellow
Cancer Biology Laboratory
Department of Biotechnology
Indian Institute of Technology Madras, India 600036
Email: satarupabando@gmail.com, satarupa.biotech@gmail.com

- **Phone +91-7557873426, +91-947400526**
- **Researchgate -** https://www.researchgate.net/profile/Satarupa_Banerjee3
- **Google Scholar -** https://scholar.google.co.in/citations?user=akNN4IwAAAAJ&hl=en
- **Publon -** https://publons.com/researcher/535973/satarupa-banerjee

Education:

- **Indian Institute of Technology Madras, India,** Institute Postdoctoral Fellow, August 2017 - Continuing
- **La Jolla Institute for Allergy and Immunology, CA, USA** Post-Doctoral Fellow, August 2016 – February 2017
- **Indian Institute of Technology, Kharagpur, India,** PhD, School of Medical Science and Technology, July 2016
- **National Institute of Technology, Durgapur, India,** M.Tech, Biotechnology, September 2011, CGPA: 8.72/10
- **The West Bengal University of Health Sciences, India (Institute of Pharmacy, Jalpaiguri),** Bachelor of Pharmacy, September 2009, 77.63 %
- **Visva- Bharati (Central University, Santiniketan, West Bengal, India**) 12th Standard (Pre Degree Examination), July 2005, 79.99 %
- **Visva- Bharati (Central University, Santiniketan, West Bengal, India)**10th Standard (School Certificate Examination), May 2003, 86.77 %

Courses taught:

- Teaching Assistant for BT6310: Cancer Biology in IIT Madras for Jan-May 2018 and Jan-May 2019
- Teaching Assistant for MM71514: Molecular Imaging in IIT Kharagpur

Awards and Achievements:

- Awarded for “Outstanding Contribution in Reviewing” from Journal of Biomedical Informatics, Elsevier in September 2018
- Dean Travel Grant for participating ACS National Meeting and Exposition in USA, 2016 by IIT Kharagpur
- Graduate Aptitude Test in Engineering (GATE) 2009 Paper: Pharmaceutical Sciences (PY) and NIPER 2009
- MHRD Fellowship for Doctoral studies 2011- 2016

- Best Poster Award in Research Scholar Day, SMST, 2016, IIT Kharagpur
- Best Poster Award in 7th International Conference on Translational Cancer Research, Chennai
- Best Paper Award in Oral Presentation Category in 18th State Science & Technology Congress, West Bengal, 2010

Certification from Coursera (Completed)

- Introduction to Genomics Technologies authorized by Johns Hopkins University, USA
- Python for Genomics Data Science authorized by Johns Hopkins University, USA
- Bioconductor for Genomics Data Science authorized by Johns Hopkins University, USA

Publications:

Publication in Journals

- **S Banerjee**, SRK Yabalooru and D Karunagaran, "Identification of mRNA and non-coding RNA hubs using network analysis and graph theory in organ tropism regulated triple negative breast cancer metastasis" Under Review in "Journal of Advanced Research, Elsevier"
- **S Banerjee** and D Karunagaran. "An Integrated Approach for Mining Precise RNA-based Cervical Cancer Staging Biomarkers" Gene, 712: 2019 (Link https://www.sciencedirect.com/science/article/pii/S0378111919306110?via%3Dihub)
- A.G. Mazumder, **S. Banerjee**, F. Zevictovich, S. Ghosh, A. Mukherjee, J. Chatterjee. "Fourier-Transform-Infrared-Spectroscopy based Metabolomic Spectral Biomarker Selection towards Optimal Diagnostic Differentiation of Diabetes with and without Retinopathy" Spectroscopy Letters 51.7 (2018): 340-349. (Link - https://www.tandfonline.com/doi/full/10.1080/00387010.2018.1471510)
- A Barik**, S Banerjee**, S Dhara and N Chakravorty, "A Reductionist Approach to Extract Robust Molecular Markers from Microarray Data Series - Isolating Markers to Track Osseointegration" Journal of Biomedical Informatics 68 (2017): 104-111. (Link - https://www.sciencedirect.com/science/article/pii/S1532046417300540?via%3Dihub)
- **S Banerjee,** A Anura, J Chakrabarty, S Sengupta, J Chatterjee. "Identification and functional assessment of novel gene sets towards better understanding of dysplasia associated oral carcinogenesis" Gene Reports 4 (2016): 131-138. (Link - https://www.sciencedirect.com/science/article/pii/S2452014416300243)
- **S Banerjee**, S Chatterjee, A Anura, J Chakrabarty, M Pal, B Ghosh, R R Paul, D Sheet, J Chatterjee. "Global Spectral and Local Molecular Connects with Optical Coherence Tomography Features to Classify Oral Lesions towards Unraveling Quantitative Imaging Bio-markers" RSC Advances 6.9 (2016): 7511-7520. (Link - https://pubs.rsc.org/en/content/articlelanding/ra/2016/c5ra24117k#!divAbstract)
- **S Banerjee**, M Pal, J Chakrabarty, C Petibois, RR Paul, A Giri, and J Chatterjee. "Fourier- transform infrared-spectroscopy based spectral-biomarker

selection towards optimum diagnostic differentiation of oral leukoplakia and cancer." Analytical and bioanalytical chemistry 407. 26 (2015): 7935- 7943. (Link -)

- **S Banerjee,** Aishwaryaprajna, D Chakraborty, A Giri, R Ghosh, B C Sarkar and J Chatterjee. “Application of Fuzzy Logic for Assessing Oral Precancer and Cancer Susceptibility", Egyptian Informatics Journal 17. 3 (2016): 251–263. (Link - https://www.sciencedirect.com/science/article/pii/S1110866516000037)
- HH Chen, V Bobroff, M Delugin, R Pineau, R Noreen, Y Seydou, **S Banerjee,** J Chatterjee, S Javerzat and C Petibois. “The future of infrared spectroscopy in biosciences: in vitro, time-resolved, and 3D” Acta Phys Pol A 129.2 (2016): 255-259. (Link - http://psjd.icm.edu.pl/psjd/element/bwmeta1.element.bwnjournal-article-appv129n226kz)
- **S Banerjee** and J Chatterjee "Molecular Pathology Signatures in Predicting Malignant Potentiality of Dysplastic Oral Pre-cancers." Springer Science Reviews 3.2 (2015): 127-136. (Link - https://link.springer.com/article/10.1007/s40362-015-0033-7)
- **S Banerjee** and J Chatterjee. "Efficient extraction strategies of tea (Camellia sinensis) biomolecules." Journal of Food Science and Technology 52.6 (2014): 3158-3168. (Link - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4444893/)
- N Sarkar, SK Ghosh, **S Banerjee** and K Aikat. "Bioethanol production from agricultural wastes: An overview." Renewable Energy 37.1 (2012): 19-27. (Link - https://www.sciencedirect.com/science/article/pii/S096014811100382X)
- **S Banerjee** and A Mitra “Changing landscape of herbal medicine: Technology attributing renaissance” International Journal of Pharmacy and Pharmaceutical Sciences, 4 (1) (2012): 47- 52. (Link - https://pdfs.semanticscholar.org/b510/1f758e70a16c1bf4ceedeaad25fa06a30933.pdf)

Book Chapter

- **S Banerjee** and J Chatterjee, “Phytochemicals, functional food and nutraceuticals for oral cancer chemoprevention” Invited Book Chapter for ‘Food Engineering Emerging Issues, Modelling and Applications’ (Nov, 2016 as part of book series on Innovations in Agricultural and Biological Engineering), Apple Academic Press, E-Book ISBN: 9781315366258, Pages 166-187. (Link - http://www.appleacademicpress.com/food-engineering-emerging-issues-modeling-and-applications/9781771883689)
- **S Banerjee**, SK Ghosh, N Sarkar, K Aikat , “Complete commercial utilization of Sugarcane: A boost to rural economy” Integrated Rural Development and Management : Issues Strategies and Policy Options, edited by Debabrata Dasgupta, Amal Kr. Mallick, Prakash Kanti Das, Amitava Dutta, Rupak Goswami and Md. Nasim Ali,: pages 172-181; Agrobios Publisher, India, 2012., ISBN: 9788177544831

Conference Paper

- **S Banerjee**, S Chatterjee, S. P. K. Karri, M Pal, R R Paul and J Chatterjee, “Multimodal Diagnostic Segregation of Oral Leukoplakia and Cancer" International Conference on Systems in Medicine and Biology (ICSMB), Kharagpur, 2016, pp. 95-97. (IEEE Xplore) (Link - https://ieeexplore.ieee.org/document/7915089)

- **S Banerjee** and J.Chatterjee, “Functional Stratification of Biomarkers Selected from Microarray Data of Oral Leukoplakia and Cancer" International Conference on Systems in Medicine and Biology (ICSMB), Kharagpur, 2016, pp. 67-70.(IEEE Xplore) (Link - https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7915096)
- D Sheet, **S Banerjee**, SPK Karri, S Bag, A Anura, A Giri, RR Paul et al. "Transfer learning of tissue photon interaction in optical coherence tomography towards in vivo histology of the oral mucosa." 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), pp. 1389-1392. (IEEE Xplore) (Link - https://ieeexplore.ieee.org/document/6868137)
- N Sarkar, SK Ghosh, **S Banerjee,** K Aikat. Effect of Temperature and Incubation Period in Enzymatic Hydrolysis of Rice Straw for the Production of Reducing sugars by locally isolated Bacillus strain,Procedeeings of National Conference on Biotechnology and the Environment organized by the Department of Biotechnology, National Institute of Technology, Durgapur, India during October 4-5, 2010.
- SK Ghosh, **S Banerjee**, N Sarkar, K Aikat, “Acid Treatment of Sugar-cane Bagasse: A critical step for Bio- ethanol production,” International Conference on Applications of Renewable and Sustainable Energy for Industry and Society organized by the Department of Physics, Osmania University, Hyderabad, India during December 16-18, 2010.
- SK Ghosh, **S Banerjee**, N Sarkar, K Aikat, “Fermentable Sugars from Sugar-cane Bagasse… Economic Management of an Agro-industrial Waste” on “8th All India People’s Technology Congress” organized by the Forum of Scientists, Engineers and Technologists (FOSET), Science City, Kolkata, India during February 11- 12, 2010.
- SK Ghosh, **S Banerjee**, K Aikat. "Detoxification of Acid Hydrolysate of Sugarcane Bagasse for the Removal of Furfural" at the National Conference on Industrial Engineering, Organized by West Bengal University of Technology, Kolkata during February 17-18, 2011.
- SK Ghosh, **S Banerjee** and K Aikat. "Comparative Study of Three Lignocellulosic Substrates for Bioethanol Production" 18th State Science & Technology Congress, West Bengal, held from 28th February, 2011 to 1st March, 2011 at the Ramakrishna Mission Ashrama, Vivekananda Centenary Hall, Narendrapur, Kolkata: 700103 (**Conferred Best Paper Award in Oral Presentation Category**)

Conference Abstracts

- S Krishna Priya, **S Banerjee**, D Karunagaran, and G. K. Suraishkumar “Understanding the Association of RNA Hubs in Fluid Shear Stress and Reactive Oxygen Species Mediated Colon Cancer Metastasis” Accepted for presentation in 8th International Conference on Translational Cancer Research to be held in Benaras Hindu University, 13-16 February, 2020.
- **S Banerjee** and D Karunagaran “Identification of Regulatory miRNAs in Cervical Cancer Metastasis” ” in Proceedings of International Conference on Emerging Areas of Biotechnology For Human Welfare And Bioentrepreneurship held in B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India. 11-12, September, 2018. pp – 62.
- SRK Yabalooru, **S Banerjee** and D Karunagaran “Resolving pathways associated in breast carcinogenesis irrespective of hormone receptor status

from functional genomics perspectives" in Proceedings of International Conference on Emerging Areas of Biotechnology for Human Welfare and Bioentrepreneurship held in B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India. 11-12, September, 2018. pp – 63.

- **S Banerjee** and D Karunagaran "Exploiting Molecular Pathways for Unravelling Breast Cancer Associated Brain Metastasis" in Proceedings of 7th International Conference on Translational Cancer Research held in Chennai, 08-11 February, 2018. pp - 147.
- C Kishore, **S Banerjee** and D Karunagaran "Molecular Pathways and Biological Processes Regulated by Menadione in Colon Cancer" in Proceedings of 7th International Conference on Translational Cancer Research held in Chennai, 08-11 February, 2018. pp - 92.
- P Sharma, **S Banerjee** and D Karunagaran "Unravelling Pathways Associated with Target miRNAs of High-Mobility-Group Proteins through the miRNA-mRNA Network" in Proceedings of 7th International Conference on Translational Cancer Research held in Chennai, 08-11 February, 2018. pp - 133**. (Best Poster Award)**
- P Yadav, **S Banerjee** and D Karunagaran "Role of micro-RNAs in the Regulation of Ferroptosis in Mammary Gland" in Proceedings of 7th International Conference on Translational Cancer Research held in Chennai, 08-11 February, 2018. pp - 131.
- **S Banerjee,** J Chakrabarty, M Pal, RR Paul, and J Chatterjee, "Fourier Transform Infrared Spectroscopic Spectral Feature Subset Selection for Optimal Diagnosis of Oral Lesion "Presented in 251st ACS National Meeting & Exposition, on March 13-17, 2016 in San Diego, CA, USA. (Link - https://apps.webofknowledge.com/full_record.do?product=WOS&search_mode=GeneralSearch&qid=1&SID=E5fPknhLR3B6fIH3uUF&page=1&doc=1)
- **S Banerjee**, D Sheet, A Giri, R Ghosh, M Pal, RR Paul, and J Chatterjee, "Optical Coherence Tomographic Attenuation Imaging Based Oral Precancer Diagnosis" Head & Neck. 37: E133, Jul 2015 (Oral Presentation in 5th World Congress of International Academy of Oral Oncology, 8-11 July 2015, Sao Paulo, Brazil) (Link - https://onlinelibrary.wiley.com/doi/abs/10.1002/hed.24138)
- **S Banerjee**, D Sheet, RK. Das, A Anura, S Bag, A Giri, RR Paul, M Pal, B Sarkar, R Ghosh, B Ghosh, and J Chatterjee, "Recognizing Pathobiological Signatures of Oral Precancers through Multimodal Imaging" Indian Journal of Cancer Supplementary, Nov 2013, Volume 50, Supplement 1, S-221. Presented in 1st Indian Cancer Congress. 21 - 24 Nov. Kempinski Ambience Hotel, Delhi, India.
- **S Banerjee**, D Sheet, A Giri, B Sarkar, R Ghosh, and J Chatterjee, "Clinicoepidemiological perspectives of oral precancers and cancers in North Bengal, India." Indian Journal of Cancer Supplementary, Nov 2013, Volume50, Supplement 1, S-158. Presented in 1st Indian Cancer Congress. 21 - 24 Nov. Kempinski Ambience Hotel, Delhi, India.

Professional Experience:

- Reviewer of 'Food Science and Biotechnology', Springer Publication, 'Journal of Biosciences', Springer Publication, 'Journal of Biomedical Informatics', 'Gene', Elsevier

- Organizing committee member of "Interdisciplinary Workshop Towards Achieving In situ Functional Histopathology" organized by School of Medical Science and Technology, Indian Institute of Technology Kharagpur, during 21-22 March 2015 (Link - http://www.facweb.iitkgp.ac.in/~debdoot/aifh2015/)

- Workshop attended on "Bioinformatics in Genomics, Proteomics and Metabolomics" organized by Department of Biotechnology, Indian Institute of Technology, Kharagpur during September 24-25, 2010

Name of Referees:

- Devarajan Karunagaran, PhD

Professor
Institute Chair Professor and Head, Department of Biotechnology
Bhupat and Jyoti Mehta School of Biosciences,
Indian Institute of Technology Madras, India - 600036, Tamilnadu
Phone: 044-22574126 (Office)
E-mail: karuna@iitm.ac.in

- Jyotirmoy Chatterjee, PhD

Professor and Head of the Department
School of Medical Science and Technology
Indian Institute of Technology Kharagpur,
West Bengal, 721302, India
Phone: 03222 282302 (Office)
Mobile: +91 9733742446
Fax: +91 3222 282221 (Office)
Email: jchatterjee@smst.iitkgp.ernet.in

- Jitamanyu Chakrabarty, Ph.D.

Associate Professor
Department of Chemistry, National Institute of Technology Durgapur
West Bengal, 713209, India
Mobile: +91 9433316694
Email: jitamanyu.chakrabarty@ch.nitdgp.ac.in; jito19@gmail.com

Anal Bioanal Chem (2015) 407:7935–7943
DOI 10.1007/s00216-015-8960-3

RESEARCH PAPER

Fourier-transform-infrared-spectroscopy based spectral-biomarker selection towards optimum diagnostic differentiation of oral leukoplakia and cancer

Satarupa Banerjee[1] · Mousumi Pal[2] · Jitamanyu Chakrabarty[3] · Cyril Petibois[4] · Ranjan Rashmi Paul[2] · Amita Giri[5] · Jyotirmoy Chatterjee[1]

Received: 31 May 2015 / Revised: 24 July 2015 / Accepted: 4 August 2015 / Published online: 5 September 2015

Abstract In search of specific label-free biomarkers for differentiation of two oral lesions, namely oral leukoplakia (OLK) and oral squamous-cell carcinoma (OSCC), Fourier-transform infrared (FTIR) spectroscopy was performed on paraffin-embedded tissue sections from 47 human subjects (eight normal (NOM), 16 OLK, and 23 OSCC). Difference between mean spectra (DBMS), Mann–Whitney's *U* test, and forward feature selection (FFS) techniques were used for optimising spectral-marker selection. Classification of diseases was performed with linear and quadratic support vector machine (SVM) at 10-fold cross-validation, using different combinations of spectral features. It was observed that six features obtained through FFS enabled differentiation of NOM and OSCC tissue (1782, 1713, 1665, 1545, 1409, and 1161 cm^{-1}) and were most significant, able to classify OLK and OSCC with 81.3 % sensitivity, 95.7 % specificity, and 89.7 % overall accuracy. The 43 spectral markers extracted through Mann–Whitney's *U* Test were the least significant when quadratic SVM was used. Considering the high sensitivity and specificity of the FFS technique, extracting only six spectral biomarkers was thus most useful for diagnosis of OLK and OSCC, and to overcome inter and intra-observer variability experienced in diagnostic best-practice histopathological procedure. By considering the biochemical assignment of these six spectral signatures, this work also revealed altered glycogen and keratin content in histological sections which could able to discriminate OLK and OSCC. The method was validated through spectral selection by the DBMS technique. Thus this method has potential for diagnostic cost minimisation for oral lesions by label-free biomarker identification.

Keywords FTIR · Oral leukoplakia · Oral squamous-cell carcinoma · Forward feature selection · Support vector machine

Electronic supplementary material The online version of this article (doi:10.1007/s00216-015-8960-3) contains supplementary material, which is available to authorized users.

✉ Satarupa Banerjee
satarupa@smst.iitkgp.ernet.in

[1] School of Medical Science and Technology, Indian Institute of Technology, Kharagpur 721302, India

[2] Department of Oral and Maxillofacial Pathology, Guru Nanak Institute of Dental Science and Research, 157/F Nilganj Road, Panihati, Kolkata 700 114, India

[3] Department of Chemistry, National Institute of Technology, Durgapur 713209, India

[4] University of Bordeaux – Inserm U1029 LAMC – Biophysics of Vascular Plasticity, 33608 Pessac, France

[5] Department of Pathology, North Bengal Medical College and Hospital, Darjeeling 734012, India

Introduction

Oral carcinogenesis is a complex multistep phenomenon. Its progression starts from benign hyperplasia and evolves through dysplasia, carcinoma in situ, and finally to oral squamous-cell carcinoma (OSCC) [1]. Oral pre-malignant disorders (PMDs) are believed to be an intermediate step of oral cancer development. Among many PMDs, oral leukoplakia (OLK) is the most prevalent, and histopathological evaluation of biopsies is the diagnostic method of choice [2]. The definition of OLK, as amended in the workshop of the WHO Collaborating Centre for Oral Cancer and Pre-cancer in 2005 by the working group, is: "The term leukoplakia should be used to recognise white plaques of questionable risk having excluded (other) known diseases or disorders that carry no

RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2016, **6**, 7511

Global spectral and local molecular connects for optical coherence tomography features to classify oral lesions towards unravelling quantitative imaging biomarkers†

Satarupa Banerjee,*[a] Swarnadip Chatterjee,[b] Anji Anura,[a] Jitamanyu Chakrabarty,[c] Mousumi Pal,[d] Bhaskar Ghosh,[e] Ranjan Rashmi Paul,[d] Debdoot Sheet[f] and Jyotirmoy Chatterjee[a]

Received 15th November 2015
Accepted 28th December 2015

DOI: 10.1039/c5ra24117k

www.rsc.org/advances

The biopsy based diagnosis of oral precancers like leukoplakia (OLK) and submucous fibrosis (OSF) as well as squamous cell carcinoma (OSCC) suffers from observer specific variability. The present work explores the utility of intensity and textural features from optical coherence tomography (OCT) images after specific feature subset selection for precise classification of oral lesions using variants of support vector machine. Concomitant application of Fourier transform infrared (FTIR) spectroscopy for endorsing global biochemical signatures, and histochemistry was performed further for value addition of the OCT findings. Immunohistochemical findings for characterization of specific local molecular alteration were also included in this. Result suggested that, OCT features could differentiate the lesions with high sensitivity and specificity. The FTIR result showed glycogen, keratin and carbohydrate related alteration in OSCC, decrease in collagen specific amino acids and skeletal muscle related proteins in OSF and distinct variation in tissue hydration status in diseases. There was also increase in keratin layer thickness in OLK due to overexpression of cytokeratin 10 in superficial layer; while in OSF, skeletal muscle was found to be replaced with dense collagen I. These disease specific alterations were assumed to be the underlying phenomenon associated with intensity and textural variations in OCT images, using which specific quantitative imaging biomarkers were proposed.

Introduction

Optical diagnostic systems like optical coherence tomography (OCT), Fourier transform infrared spectroscopy (FTIR), Raman spectroscopy, microendoscopy, and fluorescence spectroscopy are effectively emerging for non-invasive studies of pathologies, especially for characterization of pre-cancer and cancer. These techniques also help in value addition to the existing histopathological diagnostic gold standard as well as molecular pathology towards exploration of newer information.[1]

OCT, a non-invasive imaging technique, provides real-time, high-resolution, micro-architectural sub-surface images of up to 2 mm tissue depth.[2] Previous studies correlated healing progression and maturation of epithelial and sub-epithelial components considering OCT image attributes and histopathological features.[3] 'Lucidity' is the optical intensity descriptor used for interpreting OCT images. It tends to vary in different regions of layered body structure like oral mucosa, skin wounds *etc.* Since the operating principle of OCT imaging is governed by backscattering of light and exploiting a 'biological window' with minimal absorption, the changes in tissue refractive index modulate intensity characteristics.[4] Such scattering also depends upon tissue structural components, surface roughness,[5] hydration cum maturation status, nuclei size, presence of collagen fibres, keratin content,[6] tissue type[7] and membrane lipid density of cells.[8] In skin,[9] cervix[10] and oral mucosa[11] transition zones and architectural changes during disease progression can be identified by OCT. Such demarcations are possible due to differential thickness and composition of epithelial or sub-epithelial layers.[12,13] In this context, Ughi *et al.* utilized intravascular coronary OCT for differentiating normal and abnormal pathologic condition by textural image analysis.[14]

[a]*School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, India. E-mail: satarupa@smst.iitkgp.ernet.in*

[b]*Advanced Technology Development Centre, Indian Institute of Technology, Kharagpur, India*

[c]*Department of Chemistry, National Institute of Technology Durgapur, India*

[d]*Department of Oral and Maxillofacial Pathology, Guru Nanak Institute of Dental Science and Research, Kolkata, India*

[e]*Department of ENT & Head Neck Surgery, Medical College, Kolkata, India*

[f]*Department of Electrical Engineering, Indian Institute of Technology, Kharagpur, India*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5ra24117k

Journal of Biomedical Informatics 68 (2017) 104–111

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

A reductionist approach to extract robust molecular markers from microarray data series – Isolating markers to track osseointegration

Anwesha Barik, Satarupa Banerjee, Santanu Dhara, Nishant Chakravorty *

School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, West Bengal 721302, India

ARTICLE INFO

Article history:
Received 14 October 2016
Revised 3 February 2017
Accepted 10 March 2017
Available online 11 March 2017

Keywords:
Osseointegration
Microarray data
Statistical classifiers
Gene ontology
Surface texture

ABSTRACT

Complexities in the full genome expression studies hinder the extraction of tracker genes to analyze the course of biological events. In this study, we demonstrate the applications of supervised machine learning methods to reduce the irrelevance in microarray data series and thereby extract robust molecular markers to track biological processes. The methodology has been illustrated by analyzing whole genome expression studies on bone-implant integration (ossointegration). Being a biological process, osseointegration is known to leave a trail of genetic footprint during the course. In spite of existence of enormous amount of raw data in public repositories, researchers still do not have access to a panel of genes that can definitively track osseointegration. The results from our study revealed panels comprising of matrix metalloproteinases and collagen genes were able to track osseointegration on implant surfaces (MMP9 and COL1A2 on micro-textured; MMP12 and COL6A3 on superimposed nano-textured surfaces) with 100% classification accuracy, specificity and sensitivity. Further, our analysis showed the importance of the progression of the duration in establishment of the mechanical connection at bone-implant surface. The findings from this study are expected to be useful to researchers investigating osseointegration of novel implant materials especially at the early stage. The methodology demonstrated can be easily adapted by scientists in different fields to analyze large databases for other biological processes.

1. Introduction

Modern day molecular biology tools and techniques like microarray, gene expression profiling have provided researchers with a vast armory to decode the molecular mechanisms of different biological processes. Advances in medicine and related fields require an in-depth understanding of these mechanisms at cellular level to identify and characterize the disease condition. In some cases, confirmation of the progress of the improvement in the clinical condition during the course of therapy also has been made easy by these techniques. Analysis of the enormous amount of data generated through these techniques is the known to be the most time consuming task [1]. The inconvenience caused thereby raises the challenge to the analysts towards extracting the most desired information. Difficulties in finding gene annotations and relating them to literature references have made this a tedious job. As a result of this, the interpretation of the data analysis cannot be easily linked to the previous studies. Finally, the rarity of standardized protocol limits the scope of the data mining. There are very few, practically none such single tool which can perform the tasks all together, including database storage, data queries, statistical analysis, clustering, functional analysis, interrelation within the relevant cluster and interaction with public databases as well as experimental outcomes on the Internet [2]. In order to extract meaningful information from freely available datasets/data series we have designed a methodology that can be easily adapted by researchers in different fields to extract succinct information from existing datasets. The application of this methodology has been demonstrated on osseointegration – phenomenon of crucial clinical importance.

The functional success of the oral prosthetics is directly related to the extent of osseointegration of the implants. Conceptualized by Branemark, osseointegration can be described as the "direct structural and functional connection between ordered, living bone and the surface of a load-carrying implant" ensuring a long term clinical stability of the implants [3,4]. Researchers working in the field of implantology often need to validate the osseointegrative potential of materials developed when compared with the existing ones. Although histological studies on samples collected from the bone-implant interface are considered to be among the most reliable methods in experimental in vivo analysis, such techniques involve invasive techniques. Comparative expression profiling of whole genome either by microarray or selective genetic profiling

* Corresponding author.
E-mail address: nishant@smst.iitkgp.ernet.in (N. Chakravorty).

http://dx.doi.org/10.1016/j.jbi.2017.03.005
1532-0464/

Gene Reports 4 (2016) 131–138

Contents lists available at ScienceDirect

Gene Reports

journal homepage: www.elsevier.com/locate/genrep

Identification and functional assessment of novel gene sets towards better understanding of dysplasia associated oral carcinogenesis

Satarupa Banerjee [a,*], Anji Anura [a], Jitamanyu Chakrabarty [b], Sanghamitra Sengupta [c], Jyotirmoy Chatterjee [a]

[a] School of Medical Science and Technology, Indian Institute of Technology, Kharagpur 721302, India
[b] Department of Chemistry, National Institute of Technology Durgapur, India
[c] Department of Biochemistry, University of Calcutta, Kolkata, India

ARTICLE INFO

Article history:
Received 31 March 2016
Received in revised form 11 April 2016
Accepted 11 April 2016
Available online 23 April 2016

Keywords:
Oral epithelial dysplasia
Oral squamous cell carcinoma
Venn diagram
Gene sub-set selection

ABSTRACT

Oral epithelial dysplasia (OED) often precedes oral cancer. Understanding the underlying complex biological aspects of dysplasia associated oral carcinogenesis using important gene sets is thus important. Computation assisted gene set identification through different feature ranking and visualization techniques was therefore attempted in this study. Result suggested that, weighted support vector machine (SVM) could be useful for feature ranking and SVM for attribute selection. Alteration in keratinization, cell–cell communication and peptidase activity was the major affected phenomena, while extracellular matrix dynamics was also found to be hampered. During best gene subset identification, set of six genes could classify normal (NOM) and oral squamous cell carcinoma (OSCC) conditions and two sets comprising four genes in each could classify NOM and dysplastic (DYS) conditions with 100% sensitivity and specificity. A gene set, comprising 32 genes showed best efficacy of 94.12% sensitivity, 99.40% specificity and 98.89% accuracy during classification of DYS and OSCC.

1. Introduction

Oral epithelial dysplasia (OED) is often a step that precedes development of squamous cell carcinoma. It can either convert to oral squamous cell carcinoma (OSCC) or revert back to normal condition, if treated early. Till date there are no specific biomarkers which may be precisely utilized to assess malignant potentiality of oral precancers including OED. Histopathological evaluation of biopsy specimens still serves as gold standard for critical detection of grades of dysplasia and for predicting its malignant potentiality. However, the procedure lacks specificity and suffers from inter and/or intra-observer variability because of the paucity of unequivocal features of dysplasia that may be regarded as cardinal markers for accurate prediction of progression risks in oral pre-malignant disorders A recent review suggested that combination of selected biomarkers may be effective to address such problem (Banerjee and Chatterjee, 2015).

OED is a histopathological condition, where cytological and architectural characteristics of oral mucosa are altered. The role of OED in oral carcinogenesis is quite controversial. Some literature suggests that likelihood of malignant transformation of OED is significant (Al-Dakkak, 2010), while other studies have shown that there is no correlation between malignant potentiality and grade of dysplasia (Dost et al., 2014). In such circumstances, understanding the molecular progression of OED to OSCC is important and can no longer be avoided (Pitiyage et al., 2009). Semi-quantitative analysis of immunohistochemically stained tissue sections has been attempted to grade OED in precancers, (Anura et al., 2014) however, the procedures are still immature and have not yet been utilized in routine clinical practices. Comparative and quantitative assessments of histological grading and immunohistochemical expression of few key molecules to study the association between OED and OSCC were reported in few studies (Anura et al., 2014; Tabor et al., 2003). Molecular dissection of oral carcinogenesis has also been attempted through the analysis of proteome and deregulation of molecular network (Molinolo et al., 2009), but understanding the progression of OED to OSCC remains in its infancy. In silico analysis of microarray gene expression data is recently gaining interest for selection of candidate gene which may be subjected to gene ontology (GO) and functional enrichment analysis for understanding underlying molecular, biological and cellular activities of given gene sets and prioritizing candidate diagnostic indicators (Hindumathi et al., 2014).

In this study, an in-depth bioinformatic and statistical analyses of the microarray transcriptome were attempted to throw light on the process. Differentially expressed (DE) genes were primarily selected to dissect progression of OSCC through OED. Weighted support vector machine (SVM) was employed to select precise gene subset towards optimal classification of oral lesions, OED and OSCC. Venn diagram was implemented in visualization of complex association of different gene sets, to unearth their possible functional association (Kestler et al., 2005). The major aim of this cost-minimized strategy exercise is

Abbreviations: OED, oral epithelial dysplasia (OED); SVM, support vector machine; NOM, normal; OSCC, oral squamous cell carcinoma; DYS, dysplastic.

* Corresponding author. Tel.: +91 9474005265.
E-mail address: satarupa@smst.iitkgp.ernet.in (S. Banerjee).

http://dx.doi.org/10.1016/j.genrep.2016.04.007
2452-0144/

Academic Sciences International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 4, Suppl 1, 2012

Review Article

CHANGING LANDSCAPE OF HERBAL MEDICINE: TECHNOLOGY ATTRIBUTING RENAISSANCE

SATARUPA BANERJEE AND ANALAVA MITRA*

School of Medical Science and Technology, Indian Institute of Technology Kharagpur 721302 West Bengal, India.
Email: analavamitra@gmail.com, amitra@adm.iitkgp.ernet.in

Received: 20 Oct 2011, Revised and Accepted: 23 Nov 2011

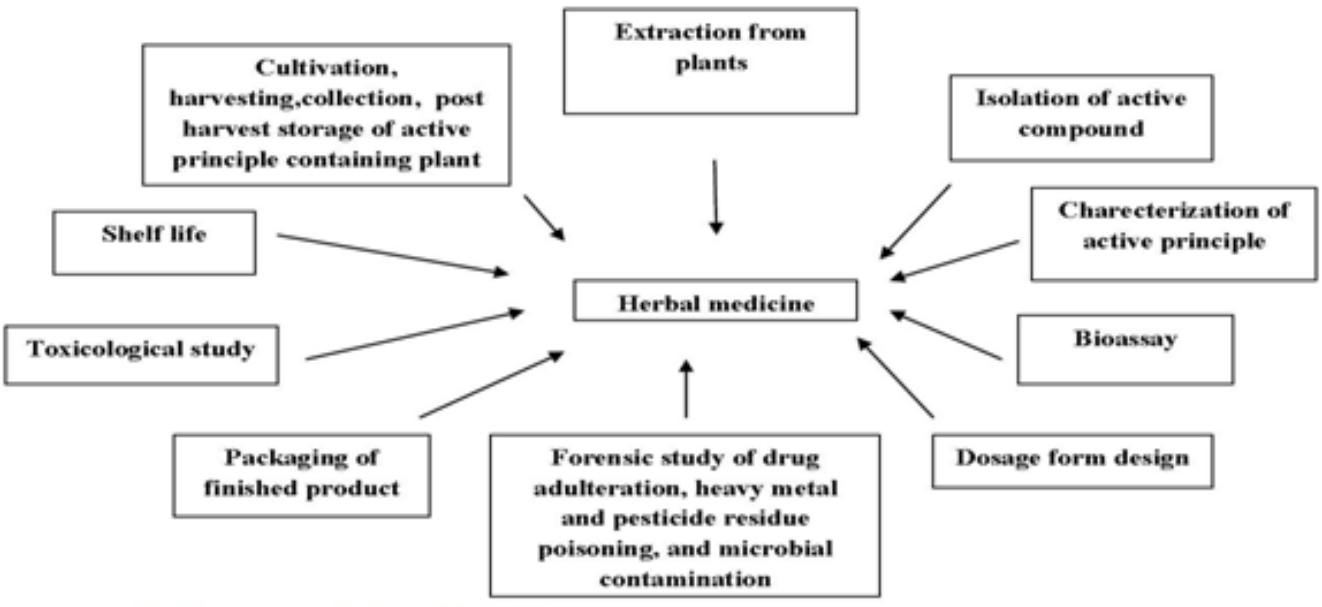
ABSTRACT

Herbal medicine is undoubtedly the oldest system of medical science in the world. Utilization of nature's wealth for health benefits and the cure, prevention and mitigation of diseases plays a big role in human civilization, with a dependency of a large number of human populations particularly in developing countries. Globalization of the local knowledge regarding the use of indigenous medicinal plants by traditional healers and localization of globally advanced technologies have boosted the growth of herbal industry and created immense global interest towards herbal medicine. The advancement in the technologies have also helped the developed countries to adopt this ancient and enriched medicinal system in a new way. An improvement in each step of herbal medicine production has recently been possible with the aid of technical developments. The increase in the demand and utilization of herbal medicine in the past few years has been increased considerably indicating herbal 'renaissance'. The current trend of utilization of these formulas after scientific researches and modern technological aspects helped in industrial growth of herbal medicine.

Keywords: Herbal Medicine.

INTRODUCTION

From the beginning of human civilization people are indebted to nature in many ways. From time immemorial plants and natural products are being used for prevention, mitigation and cure of diseases, which was the advent of primitive healthcare system. The primary health care in most of these ancient civilized societies are based on herb based medicines yet still fuzziness exists in encompassing different domains of herbal medicine as a whole (Figure 1). Contributions from ancient civilizations like Arian, Egyptian, Sumerian, Greek and others towards herbal medicine highlights conglomeration of experimental and occult knowledge specific to a particular culture. Countries like India, China, Japan, Egypt as well as Africa, Pakistan and Middle East, have their own forms of indigenous healthcare systems mostly based on herbs (Table 1). In the last part of twentieth century, Western Nations realized the importance of herbal medicine as the one that possesses maximum health benefits with minimum adverse effects and countries like USA, UK, Australia, and other European countries have accepted the medication [1, 2]. Herbal drugs are recently prepared mostly by eco-friendly processes from plants and can be defined as preparations containing active constituents of medicinal importance. It is also often called as phytomedicine/ botanical and considered as a part of alternative and complementary medicine. The traditional systems of medicine revived in all over the world in the light of modern technological aspects. Advancement in different areas of herbal research starting from extraction procedures to isolation and identification techniques, design and utilization of bio-assay for efficacy testing, dosage form design, and study of pharmacokinetic, pharmacodynamic, toxicological and pharmacological mode of action call for a healthy competition with existing classic health care systems. Simultaneously, uses of forensic studies in regulatory aspects as well as global marketing strategy are other highlighted areas of herbal medicine industry. The literature survey based study addresses these important issues on global perspective.



Proceedings of the 50th Zakopane School of Physics, Zakopane, Poland, May 18–23, 2015

The Future of Infrared Spectroscopy in Biosciences: *In Vitro*, Time-Resolved, and 3D

Hsiang-Hsin Chen[a], V. Bobroff[a], M. Delugin[a], R. Pineau[a], Razia Noreen[b], Yao Seydou[c], S. Banerjee[d], J. Chatterjee[d], S. Javerzat[a] and C. Petibois[a,*]

[a]Université de Bordeaux, Inserm U1029 LAMC, Allée Geoffroy Saint-Hillaire, Bat B2, 33600 Pessac, France

[b]Department of Applied Chemistry and Biochemistry, Govt. College, University, Faisalabad, Pakistan

[c]Centre for NanoHealth, Institute of Life Science 2 Building, Swansea University, Singleton Park, Swansea, Abertawe, SA2 8PP, UK

[d]School of Medical Science and Technology, Indian Institute of Technology, Kharagpur-721302, India

Infrared (IR) spectroscopy is at the cross-roads, with the requirement to compete with cutting-edge technologies in biosciences, mostly based on analytical performances dealing with the super-resolutions: time, lateral/spatial, and contrast. IR microscopy is diffraction limited in most cases, thus not accessing to high lateral/spatial resolutions. Additionally, it has a poor signal-to-noise ratio on a single scan, thus requiring long-lasting acquisitions that are not suitable to analyze ns-lasting biochemical events. However, it is unique because it provides a broad global chemical information of the sample contents. It is also unique because it does not require heavy sample preparation nor labeling and can be coupled to other techniques (multimodality). Finally, it is again unique because it provides quantitative measurements, thus suitable for 1D to 4D data exploitation procedures. This short review shows that IR spectroscopy will be certainly subjected to a second century of innovations, maintaining its influence in the panorama of cutting-edge analytical techniques.

DOI: 10.12693/APhysPolA.129.255

PACS: 87.10.Ca

1. Introduction

FTIR spectroscopy has been developed more than a century ago for analytical chemistry purposes, in 1893 in its principle by Nichols [1] and popularized a bit later by William Coblentz [2]. The purpose was to develop the first spectra database of molecules and to release tables of wavelengths at which various materials absorb IR light. For almost half a century [3], FTIR spectroscopy has remained a unique analytical resource to probe the structure of small molecules, but its influence in chemistry has been minored by the development of other techniques, notably nuclear magnetic resonance (NMR) and crystallography. After a century of instrumentation development, the main use of this technique has moved from molecular analyses towards the understanding of complex biological systems [4]. In the 90's, FTIR spectroscopy instrumentation focused on other applications with the release of spectrometers able to collect spectra on sample holders dedicated to quantitative, thus reproducible and comparable, measurements. The study of biofluids [5–7] has first shown that triplicate measurements allowed to obtain high reproducibility in spectra acquisitions with no more than 1% of change in absorptions on average [8, 9].

Another major advance in FTIR instrumentation for biological applications has been the development of the microscopes, as while their principle was established as early as 1949 [10], their ability to be used with both visible and IR modes [11] was found useful for bioscience. FTIR imaging systems have the advantage of a quite fast spectra acquisition since the microscope is coupled to an array of IR detectors, which may be linear (16 or 32 detectors) or focal plane (64×64 or 128×128 detectors) [11]. At this state of technological development, FTIR imaging was considered as cutting-edge technology for biosciences since the reduction of time for spectra acquisitions [12] allowed applications to large tissue areas and thus opened IR to the medical field [13]. The use of high-intensity (or high-photon flux) sources could be also considered as a major advantage to obtain high S/R levels with shorter acquisition durations [14–17]. Cell imaging applications were also demonstrated taking into account a few analytical requirements to fit analytical standards in biology [14, 16, 18–20]. However, in the current competition for proposing high-performance analytical means for biosciences, it seems that FTIR instrumentations suffer from a major lack of R&D efforts, and traditional manufacturers have not released major innovations for almost 20 years. It results that the most obvious advantages of FTIR spectroscopy and microscopy for biosciences — a global chemical characterization of unaltered samples without labeling — are now completely outreached by many other techniques, notably those based on shorter wavelengths (UV to X-ray range). In this context, several key challenges must be considered by the IR spectroscopy

*corresponding author; e-mail: cyril.petibois@u-bordeaux.fr

Springer Science Reviews
DOI 10.1007/s40362-015-0033-7

NARRATIVE STUDENT REVIEW

Molecular Pathology Signatures in Predicting Malignant Potentiality of Dysplastic Oral Pre-cancers

Satarupa Banerjee[1] · **Jyotirmoy Chatterjee**[1]

Received: 4 January 2015 / Revised: 17 August 2015 / Accepted: 27 September 2015

Abstract The role of dysplastic oral pre-cancers in oral squamous cell carcinoma development is well recognized, but the notion is not exclusive. Diagnostic gold standards in predicting malignant potentiality of such pre-cancers suffer from ambiguity due to inter- and intra-observer variability. In addressing such diagnostic challenges, combinatorial appraisement of molecular pathology attributes encompassing cancer hallmarks is thought to provide a wider analytical sense. Two major premalignant disorders, viz. oral leukoplakia and oral submucous fibrosis have been considered as candidate precursors of cancer here. This review highlights the molecular pathology signatures expressed in oral epithelial dysplasia and revisits the usefulness of combinatorial analysis of expressional pattern of existing molecular biomarkers in the context of proper selection of cardinal attributes from each cancer hallmark for better malignant potentiality assessment.

Keywords Oral epithelial dysplasia · Molecular pathology · Malignant potentiality

Introduction

Oral squamous cell carcinoma (OSCC), the sixth largest cause of death due to malignancy in the world [95], is a complex multistep phenomenon. Its progression from benign hyperplasia to dysplasia [68], to carcinoma in situ, and then into OSCC is reported [3]. Although oral epithelial dysplasia (OED) is considered to be an intermediate step for transformation of varied pre-cancerous lesions in oral cancer development, interestingly OSCC can develop from non-dysplastic lesions too [3]. Among many oral premalignant disorders (PMDs), premalignant lesion like oral leukoplakia (OLK) and premalignant conditions like oral submucous fibrosis (OSF) are commonly prevalent and histopathological evaluation of their biopsy is the diagnostic gold standard. However, this diagnostic approach suffers from non-specificity as well as intra- and inter-observer variability [96]. In spite of significant development in molecular pathology, the lack of specific molecular markers in predicting the malignant potentiality of PMDs is remarkable. Current research focuses on more selective and specific marker expression which can provide a better insight. As OED is also considered as a histopathologic marker of the malignant potentiality of PMDs [89], the molecular mechanism of dysplasia in different PMDs to assess malignant potentiality of PMDs will be reviewed in this article. The importance of a combinatorial approach will also be revisited to evaluate the expression profile of prime molecules in the context of cancer hallmarks.

Definition and Classification of OED

For OED, the initial definition provided by Pindborg (1977) and Lumermann et al. (1995) has not been accepted due to a lack of objectivity [3]. Architectural and cytological changes were considered as the major attributes of OED [54]. Grading of oral epithelial dysplasia is currently performed by interpretation of the epithelial features, like

Endorsed by Jyotirmoy Chatterjee.

✉ Satarupa Banerjee
satarupa@smst.iitkgp.ernet.in

[1] School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, Kharagpur 721302, India

Published online: 13 October 2015

Egyptian Informatics Journal (2016) 17, 251–263

Cairo University

Egyptian Informatics Journal

www.elsevier.com/locate/eij
www.sciencedirect.com

FULL-LENGTH ARTICLE

Application of fuzzy consensus for oral pre-cancer and cancer susceptibility assessment

Satarupa Banerjee [a,*], Aishwaryaprajna [b], Debjani Chakraborty [c], Amita Giri [d], Ranjan Ghosh [e], Badal Chandra Sarkar [e], Jyotirmoy Chatterjee [a]

[a] School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, 721302, India
[b] School of Mathematics, Techno India University, Kolkata 700091, India
[c] Department of Mathematics, Indian Institute of Technology, Kharagpur, 721302, India
[d] Department of Pathology, North Bengal Medical College and Hospital, Darjeeling 734432, India
[e] Department of Oral and Maxillofacial Pathology, North Bengal Dental College and Hospital, Darjeeling 734432, India

Received 1 January 2015; accepted 5 September 2015
Available online 3 March 2016

KEYWORDS

Fuzzy rulebase;
Oral health;
Oral habit;
Susceptibility assessment;
Oral pre-cancer and cancer

Abstract Health questionnaire data assessment conventionally relies upon statistical analysis in understanding disease susceptibility using discrete numbers and fails to reflect physician's perspectives and missing narratives in data, which play subtle roles in disease prediction. In addressing such limitations, the present study applies fuzzy consensus in oral health and habit questionnaire data for a selected Indian population in the context of assessing susceptibility to oral pre-cancer and cancer. Methodically collected data were initially divided into age based small subgroups and fuzzy membership function was assigned to each. The methodology further proposed the susceptibility to oral precancers (viz. leukoplakia, oral submucous fibrosis) and squamous cell carcinoma in patients considering a fuzzy rulebase through *If-Then* rules with certain conditions. Incorporation of similarity measures using the Jaccard index was used during conversion into the linguistic output of fuzzy set to predict the disease outcome in a more accurate manner and associated condition of the relevant features. It is also expected that this analytical approach will be effective in devising strategies for policy making through real-life questionnaire data handling.

* Corresponding author. Tel.: +91 3222 281225; fax: +91 3222 282221.
E-mail addresses: satarupa@smst.iitkgp.ernet.in, satarupabando@gmail.com (S. Banerjee).

Peer review under responsibility of Faculty of Computers and Information, Cairo University.

Production and hosting by Elsevier

1. Introduction

India experiences one of the highest incidence rates of oral cancer globally [1]. Oral cancer is the leading cancer type in men and the third most common cancer in women [2]. In India, oral cancer is usually detected at advanced stages and the five year survival rate for advanced oral cancer is very low [3], posing an important public health challenge. Hence the early detection of

http://dx.doi.org/10.1016/j.eij.2015.09.005

J Food Sci Technol (June 2015) 52(6):3158–3168
DOI 10.1007/s13197-014-1487-3

REVIEW

Efficient extraction strategies of tea (*Camellia sinensis*) biomolecules

Satarupa Banerjee · Jyotirmoy Chatterjee

Revised: 27 May 2014 / Accepted: 15 July 2014 / Published online: 6 August 2014

Abstract Tea is a popular daily beverage worldwide. Modulation and modifications of its basic components like catechins, alkaloids, proteins and carbohydrate during fermentation or extraction process changes organoleptic, gustatory and medicinal properties of tea. Through these processes increase or decrease in yield of desired components are evident. Considering the varied impacts of parameters in tea production, storage and processes that affect the yield, extraction of tea biomolecules at optimized condition is thought to be challenging. Implementation of technological advancements in green chemistry approaches can minimize the deviation retaining maximum qualitative properties in environment friendly way. Existed extraction processes with optimization parameters of tea have been discussed in this paper including its prospects and limitations. This exhaustive review of various extraction parameters, decaffeination process of tea and large scale cost effective isolation of tea components with aid of modern technology can assist people to choose extraction condition of tea according to necessity.

Keywords Tea · Biomolecules · Extraction · Optimization

Introduction

Tea is the mostly used daily beverage throughout the world, with estimated daily consumption of more than 3 billion cups (Chen and Zhou 2005). It is valued due to potential health benefits confirmed with preclinical and epidemiological studies, its aroma content, and cultural association. There is increasing demand of tea extract and isolated tea biomolecules in pharmaceutical and food industries as natural antioxidant and for other uses. Variation of processing technique produces varied tea that are of many types: Green tea, Black tea, White tea, Yellow tea, Dark tea, Pu'erh tea and Oolong tea. Mostly each component of any variety of tea is known for some amount of bioactivity or sensory attributes. Tea biomolecules mainly consists of non protein amino acid theanine, free sugars (Unachukwu et al. 2010), methylxanthine or purine alkaloid like caffeine, theobromine, theophylline and theacrine, phenolic acids like gallic acid, and eight other catechins; (+)-catechin (C), (−)-epicatechin (EC), (−)-gallocatechin (GC), (−)-epigallocatechin (EGC), (−)-catechin gallate (CG), (−)-gallocatechin gallate (GCG), (−)-epicatechin gallate (ECG) and (−)-epigallocatechin gallate (EGCG) (Peng et al. 2008). (−)-epicatechin (EC), (−)-epigallocatechin (EGC), (−)-epicatechin-3-gallate (ECG), and (−)-epigallocatechin- 3-gallate (EGCG) are the main catechin of green tea. Cumulatively they are often called as polyphenols. Though EGCG is the major tea catechin with most anticarcinogenic property, EGC has highest antioxidant efficacy followed by EGCG, EC, and ECG respectively (Lu and Chen 2008).Alkaloid molecules are responsible for stimulant activity of tea (Horžić et al. 2012). Catechin and caffeine content and sensory attributes determines the quality of tea (Choung and Lee 2011). Nowadays decaffeinated tea is more preferred because caffeine causes irritation in gastrointestinal tract, sleeplessness, cerebral cortex stimulation and excites central nervous system in people (Ye et al. 2007). Still awareness of the antioxidant properties and other health benefit like anti carcinogenic and chemopreventive effect (Fujiki 2005; Chen and Dou 2008; George et al. 2008), anti-hyperglycemic effect (Gornes et al. 1995), anti-obesity effects (Lin and Lin-Shiau 2006), anti diabetic (Sabu et al. 2002), anti ulcer effects (Maity et al. 1995), in cardiovascular diseases (Deka and Vita 2011), anti-arthritis effects (Katiyar and Raman 2011), and anxiolytic effect (Vignes et al. 2006) made this beverage popular

S. Banerjee (✉) · J. Chatterjee
School of Medical Science and Technology, Indian Institute of Technology, Kharagpur 721302, West Bengal, India
e-mail: satarupa@smst.iitkgp.ernet.in

SPECTROSCOPY LETTERS
https://doi.org/10.1080/00387010.2018.1471510

Taylor & Francis
Taylor & Francis Group

Fourier-transform-infrared-spectroscopy based metabolomic spectral biomarker selection towards optimal diagnostic differentiation of diabetes with and without retinopathy

A. G. Mazumder[a,b], S. Banerjee[a], F. Zevictovich[c], S. Ghosh[d], A. Mukherjee[e] and J. Chatterjee[a]

[a]School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, West Bengal, India; [b]Johns Hopkins University School of Medicine, Baltimore, MD, USA; [c]Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY, USA; [d]Department of Ophthalmology Calcutta National Medical College and Hospital, Kolkata, West Bengal, India; [e]Department of Electrical Engineering, Indian Institute of Technology, Kharagpur, West Bengal, India

ABSTRACT

The purpose of this study is to identify Fourier transform infrared (FTIR) spectroscopy-based serum metabolomic spectral biomarkers using chemometrics for the diagnosis of Diabetic Retinopathy. FTIR spectroscopy was performed on 85 human serum samples [30 type 2 diabetes patients each without retinopathy and with retinopathy along with 25 normal healthy individuals as control]. Difference between mean spectra (DBMS), forward feature selection (FFS), and Mann–Whitney's *U* tests were applied for spectral biomarker selection. Classification of disease conditions was achieved using analysis of different combinations of spectral features with linear, quadratic, and cubic Support Vector Machine at 10-fold cross validation. Twelve spectral signatures extracted by FFS could differentiate diabetes and diabetic retinopathy with 90% sensitivity, 92.7% specificity, and 90.5% overall accuracy. Two peaks (1042, 1114.18 cm^{-1}) were associated with carbohydrate and polysaccharide content and five peaks (1114.18, 1165, 1211.18, 1402.70, 1451.14, 1657 cm^{-1}) represented aberrations in total lipid content. Four peaks (1114.18, 1117, 1147, 1165 cm^{-1}) were associated with protein phosphorylation and three peaks (1527, 1544.71, 1591.10 cm^{-1}) with Amide II group. Again, lipidic signatures were strongly corroborated with glycosylated hemoglobin levels in diabetic retinopathy and diabetic subjects. Spectral signatures also revealed an elevated level of β-sheet containing proteins in serum in diabetic retinopathy condition. The method was validated through spectral biomarker selection by the DBMS technique. Thus, this method has the capability of diagnostic cost minimization for detection of diabetic retinopathy by label-free spectral biomarker identification.

ARTICLE HISTORY
Received 13 August 2017
Accepted 27 April 2018

KEYWORDS
Fourier transform infrared spectroscopy; spectral biomarker selection; diabetic retinopathy; diagnostic discrimination; chemometric methods

Introduction

Diabetic retinopathy is a microcirculatory complication of the eye which occurs due to prolonged hyperglycemia. It has been predicted that global prevalence of diabetes will increase from 366 million in 2011 to 552 million in 2030,[1] and the number of people with diabetic retinopathy will raise from 126.6 million in 2010 to 191.0 million by 2030.[2] These data highlight the substantial worldwide public health burden from diabetic retinopathy and the importance of identifying risk factors.[3] The level of Glycosylated hemoglobin (HbA1c), duration of diabetes, age of the patient has direct effect on the evolution of diabetic retinopathy.[4] The other major risk factors in diabetes like blood pressure, dyslipidemia, proteinuria[5] can also augment the symptoms of diabetic retinopathy. There are studies which have proved that high diastolic blood pressure among young people and the high systolic blood pressure among elderly individuals worsen the retinopathy.[6] An increased level of cholesterol and/or of triglycerides (dyslipidemia) is also an important risk factor for diabetic retinopathy.[7] So, specific metabolomic spectral biomarkers are essential for identifying diabetic retinopathy, since these metabolic dysfunctions may persist for several years before the manifestation of diabetes clinically and its sequel (*viz.* diabetic retinopathy). Metabolomics is the profiling of metabolites in biofluids, cells, and tissues and it is routinely applied as a tool for biomarker discovery.[8] Since the metabolome is downstream from the proteome and transcriptome,[9] it represents a more elusive level of organization than the proteome or transcriptome for comprehending a complex biological system. A strong advantage of metabolomics is the ability to uncover novel and potentially relevant metabolites which can be the basis of therapeutic approaches or prognostic indicators.[10] Identification of diabetic retinopathy specific markers is necessary since metabolic spectral signatures always overlap between diabetes and diabetic retinopathy,[11] before

CONTACT Arpan Guha Mazumder arpangm@smst.iitkgp.ernet.in School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, West Bengal 721302, India; Jyotirmoy Chatterjee jchatterjee@smst.iitkgp.ernet.in School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, West Bengal 721302, India.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lstl.