Review

Ayurgenomics for stratified medicine: TRISUTRA consortium initiative across ethnically and geographically diverse Indian populations

Bhavana Prasher a,b,d,*, Binuja Varma a, Arvind Kumar a, Bharat Krishna Khuntia a, Rajesh Pandey a, Ankita Narang a, Pradeep Tiwari a,d, Rintu Kumut c,d, Debleena Guin b, Ritushree Kukreti b, Debasis Dash a,c,d, TRISUTRA Ayurgenomics Consortium, Mitali Mukerji a,b,d,*

a CSIR Ayurgenomics Unit-TRISUTRA, CSIR-Institute of Genomics and Integrative Biology, Mathura Road, New Delhi 110020, India
b Genomics and Molecular Medicine & CSIR-Institute of Genomics and Integrative Biology, Mathura Road, New Delhi 110020, India
c G.N.Ramachandran Knowledge Centre for Genome Informatics, CSIR-Institute of Genomics and Integrative Biology, Mathura Road, New Delhi 110020, India
d Academy of Scientific & Innovative research, CSIR-IGIB, Delhi, India

ARTICLE INFO

Article history:
Received 18 February 2016
Received in revised form 2 July 2016
Accepted 21 July 2016

Keywords:
Ayurgenomics
Prakriti
P4 medicine
Stratified medicine
Tridosha
Drug discovery and development
Pharmacogenomics

ABSTRACT

Background: Genetic differences in the target proteins, metabolizing enzymes and transporters that contribute to inter-individual differences in drug response are not integrated in contemporary drug development programs. Ayurveda, that has propelled many drug discovery programs albeit for the search of new chemical entities incorporates inter-individual variability “Prakriti” in development and administration of drug in an individualized manner. Prakriti of an individual largely determines responsiveness to external environment including drugs as well as susceptibility to diseases. Prakriti has also been shown to have molecular and genomic correlates. We highlight how integration of Prakriti concepts can augment the efficiency of drug discovery and development programs through a unique initiative of Ayurgenomics TRISUTRA consortium.

Methods: Five aspects that have been carried out are (1) analysis of variability in FDA approved pharmacogenomics genes/SNPs in exomes of 72 healthy individuals including predominant Prakriti types and matched controls from a North Indian Indo-European cohort (2) establishment of a consortium network and development of five genetically homogeneous cohorts from diverse ethnic and geo-climatic background (3) identification of parameters and development of uniform standard protocols for objective assessment of Prakriti types (4) development of protocols for Prakriti evaluation and its application in more than 7500 individuals in the five cohorts (5) Development of data and sample repository and integrative omics pipelines for identification of genomic correlates.

Results: Highlight of the study are (1) Exome sequencing revealed significant differences between Prakriti types in 28 SNPs of 11 FDA approved genes of pharmacogenomics relevance viz. CYP2C19, CYP2B6, ESR1, F2, PGR, HLA-B, HLA-DQA1, HLA-DB1, LDLR, CTR, CPS1. These variations are polymorphic in diverse Indian and world populations included in 1000 genomes project. (2) Based on the phenotypic attributes of Prakriti we identified anthropometry for anatomical features, biophysical parameters for skin types, HRV for autonomic function tests, spirometry for vital capacity and gustometry for taste thresholds as objective parameters. (3) Comparison of Prakriti phenotypes across different ethnic, age and gender groups led to identification of invariant features as well as some that require weighted considerations across the cohorts.

Conclusion: Considering the molecular and genomics differences underlying Prakriti and relevance in disease pharmacogenomics studies, this novel integrative platform would help in identification of differently susceptible and drug responsive population. Additionally, integrated analysis of phenomic and genomic variations would not only allow identification of clinical and genomic markers of Prakriti for application in personalized medicine but also its integration in drug discovery and development programs.

© 2016 Elsevier Ireland Ltd. All rights reserved.

* Correspondence to: CSIR Ayurgenomics Unit- TRISUTRA, CSIR- Institute of Genomics & Integrative Biology, Sukhdev Vihar, Mathura Road, New Delhi 110025, India.
** Correspondence to: Genomics and Molecular Medicine & CSIR Ayurgenomics, Unit- TRISUTRA, CSIR-Institute of Genomics and Integrative Biology, Sukhdev Vihar, Mathura Road, New Delhi 110 025, India.
E-mail addresses: bhavana.p@igib.res.in (B. Prasher), mitali@igib.res.in (M. Mukerji).

http://dx.doi.org/10.1016/j.jep.2016.07.063
0378-8741/© 2016 Elsevier Ireland Ltd. All rights reserved.

Contents

1. Introduction ............................................................ 2
2. Setting up an interdisciplinary platform in the TRISUTRA consortium: collaborative research networks ................................................. 3
3. Variability in pharmacogenomics markers amongst Prakriti groups ................................................................. 3
4. Phenome stratification using Prakriti and objective parameters ......... 4
   4.1. Anatomical feature assessments through anthropometry .............. 8
   4.2. Objective assessment of skin phenotypes .................................. 10
   4.3. Assessment of vital capacity through spirometry ...................... 10
   4.4. Heart rate variability for study of physiological attributes .......... 10
   4.5. Taste threshold assessment through electrogustometry ............ 10
5. Prakriti stratification of population in diverse cohorts of genetically homogeneous populations ............... 12
   5.1. Population identification ............................................... 12
   5.2. Preparatory phase ..................................................... 12
   5.3. Community engagement and subject recruitment .................... 12
   5.4. Prakriti screening ...................................................... 12
   5.5. Detailed phenotypic assessment using Prakriti methods and objective parameters .................................................. 12
6. Development of a biological sample repository ........................................... 13
7. Genomics and molecular analysis of individuals classified by their Prakriti types in healthy and diseased states ......................... 14
   7.1. Biochemical profiling ................................................. 14
   7.2. Dissecting phenome through OMICS ................................. 14
   7.2.1. Establishing the genetic homogeneity of study populations ..... 14
   7.3. Integrated data analysis and discovery of predictive markers ....... 14
   7.4. Validation of leads in model system .................................. 15
   7.5. Development of data-analytic platform for Ayurgenomics ......... 15
   7.6. Data sharing and attribution policy ................................... 15
8. Conclusion ................................................................. 17
Acknowledgment ........................................................................ 17
References .............................................................................. 17

1. Introduction

There has been a global increase in prevalence of chronic and complex diseases with many life-style disorders assuming near epidemic proportions. Also, slow development of new drugs has led to an unmet medical need with many of the drugs being withdrawn because of frequent side-effects. In addition, there is an enormous variability in success of treatment with respect to disease and individuals. Majority of chronic diseases require lifetime medications and in many cases such as cancer and infectious diseases, resistance to drugs is a common problem. Most of the diseases are multifactorial involving complex interplay of a network of genes and non-genetic environmental factors (Baffy and Loscalzo, 2014; Joh et al., 2007). It is being realized that we need to evolve a systems’ based approach for comprehensive understanding of biology and move towards a network approach in medicine (Auffray et al., 2010, 2009; Barabasi et al., 2011; Gibson and Visscher, 2013). With the advent of genomics, drug discovery and development program are also integrating understanding of disease biology in target identification and also aspire to identify responder populations (Beitelshes et al., 2015; Phillips et al., 2001; Suthram et al., 2010). Some of the newer generation drug development efforts have also started to incorporate genomic variation information from disease related extreme phenotypes (Harper et al., 2015). For instance loss of function mutation in PCSK9 has been linked to lower LDL-C levels (Harper et al., 2015). This has led to the development of multiple PCSK9 inhibitors like evolocumab, alirocumab for management of hypercholesterolemia.

Ayurveda since centuries has been a source of natural or plant based medicine in many drug discovery programs albeit for search of new chemical entities (Garodia et al., 2007; Holland, 1994; Patwardhan, 2005; Patwardhan and Mashelkar, 2009; Singh et al., 2007). In addition to drug molecules, systematic exploration of biological effect of herbs has also provided some unique insights for example reserpine from the herb “Sarpagandha” has been one of the most prominent drugs which led to the dissection of entire biology of dopaminergic pathways and facilitated the merger of neuro-chemistry and psychopharmacology (Jain and Murthy, 2009). Efficacy and potential of Ayurvedic medicines is also evident from many recent scientific publications for example study on Ashwagandha (Withania somnifera) that lead to the discovery of a novel therapeutic strategy for Alzheimers disease reversal (Sehgal et al., 2012).

Ayurveda, apart from the therapeutic potential also has a predictive, preventive and personalized approach to health and management of disease which has been extensively documented in original texts of Caraka and Sushruta Samhita (Sharma, 1981, 1999). This potential has not been harnessed effectively in any drug discovery programs.

The contemporary thinking of network medicine resonates with the concepts of “trisutra” that deals with the translational aspects of Ayurveda (Prasher et al., 2008; Sethi et al., 2011). This is achieved through integration of the three axes of cause (hetu), feature (linga) and therapeutics (aushadha) for development and administration of drug in an individualized manner (Prasher et al., 2016). The methods take into account baseline thresholds of an individual, his/her perturbation from the baseline thresholds in disease and selection of an appropriate drug that can restore homeostasis (Prasher et al., 2016). Our group has provided the first evidence of molecular and genome wide expression correlates of extreme Prakriti types using an Ayurgenomics approach (Prasher et al., 2008). We could also demonstrate how stratification of individuals could help identify molecular axes like hypoxia and hemostasis to be varying amongst healthy individuals that could
modulate phenotypic outcome in hypoxic conditions for example in asthma, high altitude pulmonary edema as well as thrombosis (Aggarwal et al., 2015, 2010b; Ahmad et al., 2012). The hypoxia axis that is modulated by variation in EGLN1 in a Prakriti specific manner is a prominent drug target in diseases where hypoxia is a cause or consequence (Haase, 2010; Harten et al., 2010; Lee et al., 2008; Natarajan et al., 2006; Sen Banerjee et al., 2012; Soni, 2014). Therefore, inherent variability in this gene among individuals can determine differential therapeutic outcomes. Evidence for molecular correlates of Prakriti has also been independently reported in other studies (Bhushan et al., 2005; Govindaraj et al., 2015; Juyal et al., 2012; Rotti et al., 2015).

These initial observations provided an impetus for the establishment of the TRISUTRA (Translational Research and Innovative Science Through Ayurgenomics) consortium with the primary objective of identification of disease genes and predictive markers that could also augment drug discovery programs (www.trisutra.in). One of the primary activities of the consortium is to integrate the trisutra principles of Ayurveda into contemporary practice of preventive and personalized medicine and drug development programs. This requires the development of a platform that could enable cross-talk between the contrasting disciplines through a unified language. We report five aspects of the consortium initiative (1) analysis of variability in FDA approved pharmacogenomics genes/SNPs in healthy individuals of predominant Prakriti types through exome sequencing in a North Indian cohort (2) establishment of a consortium network and development of five genetically homogeneous cohorts from populations of diverse ethnic and geo-climatic backgrounds (3) development of protocols for Prakriti evaluation and its application in more than 7500 individuals in the five cohorts (4) identification of parameters and development of uniform standard protocols for objective assessment of Prakriti types (5) Development of data and sample repository and integrative omics pipelines for identification of genomic correlates.

2. Setting up an interdisciplinary platform in the TRISUTRA consortium: collaborative research networks

CSIR-TRISUTRA consortium is developed with an aim to provide scientific credece & global acceptability to Ayurveda, new leads in genomics and molecular medicine and create inter disciplinary expertize in Ayurgenomics. This is housed in the Institute of Genomics and Integrative Biology (IGIB), a constituent laboratory of Council of Scientific and Industrial Research (CSIR), as a hub with collaborative centres as nodes spread across the country. CSIR-IGIB is involved in research in the areas of predictive medicine of common and complex disorders through population, disease and functional genomics studies. It has established a high-throughput OMICS, a next generation sequencing facility, computational and system biology programs along with infrastructure for model systems such as yeast, Drosophila and mouse for follow-up of genetic leads and functional validation studies. The major focus areas include neuro-psychiatric, respiratory, skin diseases and metabolic disorders. The institute through CSIR led Indian Genome Variation Consortium efforts has also provided the first comprehensive genetic map of the Indian population with a focus for disease genetic studies (Consortium, 2008). Applicability of this data has been realized in many disease genomics studies including Ayurgenomics (Aggarwal et al., 2015, 2010a; Gautam et al., 2015; Giri et al., 2014; Gupta et al., 2011; Jha et al., 2012; Kanchan et al., 2015; Talwar et al., 2016). This provides a natural ecosystem for CSIR-TRISUTRA Ayurgenomics research.

The TRISUTRA Consortium (would be referred as TRISUTRA) includes an interdisciplinary team of experts from Ayurveda, genomics, modern medicine, public healthy cohorts and computational biology team with the national coordination and networking provided by CSIR-IGIB (Table 1). In order for Prakriti based phenotypic stratification to be globally applicable we felt the need to undertake studies in diverse ethnic and geographical populations. For this purpose we have set up Ayurgenomics cohorts with collaborators. Further for validation of these leads in disease cohorts we have established linkages with premier tertiary referral centres of modern medicine institutes like All India Institute of Medical Sciences, Delhi and National Institute of Mental Health, Bengaluru.

3. Variability in pharmacogenomics markers amongst Prakriti groups

Although extensive variations at genetic level are reported in genes of pharmacogenomics relevance such as phase I and II metabolizing enzymes, transporters and nuclear receptors, only a limited set of variants have a reported association with drug response (Goldstein et al., 2003). Amongst these variations a few are also approved by the US Food and Drug Administration (US-FDA, http://www.fda.gov/Drugs). A recent study on 146 pharmacogenes in the 1000 Genomes Project data (n=1092 individuals) and the Exome Sequencing Project (ESP; n=6503 individuals) demonstrate that these clinically important genes are highly variable and differ considerably between populations (Kozyra et al., 2016). The study also highlighted potential priority genes for population-adjusted genetic profiling strategies and the need to generate this baseline information from diverse populations. Despite this, even within a population there is a substantial difference in drug response which indicates the need for identification of subpopulation that would respond uniformly to specific treatments.

We wanted to explore whether integrating Prakriti methods might allow identification of subpopulations within genetically homogeneous groups that might have different therapeutic response. An in-house exome sequence data from 72 individuals of extreme predominant Prakriti types comprising 18 of each Vata, Pitta and Kapha with equal number of males and females along with 18 individuals as background controls from the same ethnicity was available. Briefly, exome enrichment of samples were carried out using TrueSeq Expanded Exome enrichment protocols and sequencing was performed on HiSeq 2000 using TrueSeq SBS and 100 bps paired end chemistry. Variants were identified using GATK Unified genotyper and annotated using SNPEFF (v2.0.5) and Seattle Seq Annotation 138 server. Using Fisher’s exact test through PLINK (v 1.7) differentiating variants between two Prakriti groups were identified and variants with a p-value < 0.05 in pairwise comparisons were considered as significant. In addition, permutation analysis was done by randomly shuffling the Prakriti labels ~ 50,000 times per SNP. Those SNPs were retained whose p-values were in the lower 5% distribution of p-values of the permuted set for each SNP.

From the list of differentiating SNPs (the details of which are being communicated elsewhere) we looked at overlap of genes with the 47 genes that are US-FDA approved pharmacogenomic biomarkers in drug labeling. These pharmacogenes have been recommended by USFDA to be included in dosing strategies (Kitzmiller et al., 2011). We identified 28 SNPs mapping to 11 genes that differ significantly between Prakriti types, with some having reported associations. All the exonic variants are polymorphic in the 1000 genomes representative populations that include African (AFR), East Asians (EAS), Americans (AM), Europeans (EUR) and South Asians (SAS). Amongst the South Asians there are four populations from the Indian subcontinent that are also represented in the Ayurgenomics cohorts (described below). These include...
Bengali from Bangladesh (BEB), Punjabi from Lahore, Pakistan (PIL), Gujarati Indian from Houston Texas (GHI) and Indian Telugu in UK (TU). Details of the genes that differ between the Prakriti, their function, pharmacogenomics relevance and frequency amongst Prakriti types and in different populations is provided in the Table 2a. As is evident most of the variations are reported in genes that metabolize a large fraction of drugs such as CYP2C19 and CYP2B6, receptor genes like ESR1, PGR, LDLR whose variability determined outcomes in different therapies or important therapeutic targets such as CFTR, F2 etc. (Table 2b).

In some cases, consequence of genetic differences in responsiveness resonates with phenotypic attributes of Prakriti. For examples, Pitta Prakriti individuals are more prone to bleeding and the frequency of the alternate allele for F2 variant that confers bleeding risk in response to warfarin (Botton et al., 2011; Shikata et al., 2004) is significantly higher in Pitta compared to Kapha as well as background population (Aggarwal et al., 2015; Prasher et al., 2008). This indicates that there is a decreased dose requirement of warfarin in Pitta individuals. We also observed a significant difference between Vata and Kapha Prakriti with respect to CYP2C19. The CYP2C19*2 allele which is linked to poor metabolizer activity has significantly higher frequency in Vata compared to Kapha (Scott et al., 2013). This suggests that in Vata individuals there might be an elevated risk due to under-dosing where the substrate is a pro-drug such as tamoxifen, clopidogrel or toxicity and life threatening effects due to reduced capacity to eliminate the drug such as anxiolytic sedatives (Li-Wan-Po et al., 2010).

We also observed a non-synonymous exonic variation, rs1137101 A/G, (Gln223Arg) in the LEPR gene that been reported to be associated with adverse effects of antipsychotic drugs and anti-epileptic drug (valproic acid) in terms of bodyweight gain (Lee and Bishoph, 2011; Li et al., 2015). The frequencies of the SNPs of LEPR gene is in the direction of risk to Kapha and/or protection in Pitta for adiposity, CVD, and metabolic syndrome. This SNP also demonstrates significant variability amongst the 1000 genomes population. Earlier, we have reported significant genetic difference in another SNP rs1171271 of LEPR the leptin receptor (LEPR) gene between the Prakriti types. (Aggarwal et al., 2015) (Fig. 1). This variation in LEPR that differentiates between the Prakriti types has been linked with obesity, insulin resistance, diabetes like traits and in a recent GWAS study also to levels of plasma soluble leptin receptor (Sun et al., 2010). We observed that the protective C allele of rs1171271 (LD with GWAS SNP) also contributes to within population differences where, Pitta has 27% while IE and Kapha has 0% frequency of the protective CC genotype. Such signals do not get highlighted when populations are not stratified by phenotypes and the SNPs assume an average frequency of the background. This gene has also been identified as a potential candidate under selection and displays significant differences in allele frequencies across Indian as well as East Asian populations (Abecasis et al., 2012).

Table 1: TRISUTRA Collaborating Network highlighting the interdisciplinary expertise and activities that have been undertaken with each partner.

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Collaborating centres</th>
<th>Outreach/Expertise</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayurveda</td>
<td>Ji Roy State Ayurvedic Medical College and Hospital, Kolkata, West Bengal (Eastern India)</td>
<td>Education and research Institution of Ayurveda</td>
<td>Establishment of Ayurgenomics cohorts</td>
</tr>
<tr>
<td></td>
<td>Institute of Postgraduate Training and Research in Ayurveda, Jannagar, Gujam (Western India)</td>
<td>Outreach in geographically and ethnically diverse populations</td>
<td>Stratification of healthy and diseased individuals</td>
</tr>
<tr>
<td></td>
<td>Choudhary Brahman Prakash Ayurveda Charak Sanshthan, Najafgarh, Delhi (Northern India)</td>
<td>Clinical Resources for common and complex diseases such as obesity, diabetes, asthma, COPD, skin diseases like psoriasis, neurodegenerative disease etc.</td>
<td>Ayurveda interventions</td>
</tr>
<tr>
<td></td>
<td>KLE University’s BMK Ayurveda Maha-vidyalya, Belgaum, Karnataka (Southern India)</td>
<td>Clinical set up for unique Ayurvedic treatments like panchkarma, leech therapy.</td>
<td></td>
</tr>
<tr>
<td>Modern Medicine</td>
<td>All India Institute of Medical Sciences, New Delhi</td>
<td>Premier Tertiary referral centre with clinical outreach in wide spectrum of disorders related to neurodegenerative and metabolic conditions</td>
<td></td>
</tr>
<tr>
<td>Public Health</td>
<td>KEMHRC-Vadu, Pune, Maharashtra</td>
<td>Public health cohort set up for prospective studies</td>
<td>Establishment of cohort for longitudinal follow up studies</td>
</tr>
<tr>
<td>Genomics</td>
<td>Ganit Labs, Institute of Bioinformatics and Applied Biotechnology, Bengaluru, Karnataka</td>
<td>Epigenomics</td>
<td>Involvement of epigenetics</td>
</tr>
<tr>
<td></td>
<td>Maharshi Dayanand University, Rohtak, Haryana</td>
<td>Microbial community studies and Functional Metagenomics</td>
<td>Microbiome in health, disease and Ayurveda interventions</td>
</tr>
<tr>
<td>Cell Biology</td>
<td>National Centre for Biological Sciences, Bengaluru, Karnataka</td>
<td>State of art facility for conducting studies on LCLs and stem cell biology</td>
<td>Development of LCL repository of subjects from Ayurgenomics cohorts for functional studies</td>
</tr>
<tr>
<td>Data Analytics</td>
<td>Indian Institute of Technology Mandi, Himachal Pradesh</td>
<td>Metagenomic Analysis</td>
<td>Development of methods and analysis of multi-omics data</td>
</tr>
<tr>
<td></td>
<td>Indian Statistical Institute, Kolkata, West Bengal</td>
<td>Genetic Analysis of multifactorial traits and multivariate phenotypes</td>
<td>Integrative analysis</td>
</tr>
<tr>
<td></td>
<td>Georgia Institute of Technology, Atlanta, GA, USA</td>
<td>Population and Integrative Genomics</td>
<td></td>
</tr>
</tbody>
</table>

4. Phenome stratification using Prakriti and objective parameters

We have earlier developed a questionnaire for objective assessment of Prakriti on the basis of Ayurveda descriptions (Prasher et al., 2008). In this a consideration is given to a large number of phenotypic attributes which include examination of anatomical features, physiological attributes, mental and physical attributes and sensory perceptions (Prasher et al., 2016). This involves visual examination as well as response to questions and clinical history. The attributes that are visually examined also differ between ethnic and geographical population, across age groups, gender and
Table 2a

Patterns of distribution of differentiating SNPs associated with FDA approved pharmacogenomics markers identified through exome sequencing analysis of predominant Prakriti individuals. Frequency of the alternate allele in different Prakriti groups as well as the 1000 genomes population is provided.

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Gene (class)</th>
<th>SNP/Function</th>
<th>Ref Allele</th>
<th>Alternate Allele</th>
<th>Background_freq_pooled</th>
<th>V_freq</th>
<th>P_freq</th>
<th>K_freq</th>
<th>Fischers_P_Value</th>
<th>EAS</th>
<th>AFR</th>
<th>EUR</th>
<th>SAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CFTR</td>
<td>rs1042180 (3’ UTR variant)</td>
<td>C</td>
<td>T</td>
<td>0.37</td>
<td>0.06</td>
<td>0.17</td>
<td>0.41</td>
<td>0.01, 0.032</td>
<td>0.02</td>
<td>0.25</td>
<td>0.25</td>
<td>0.23</td>
</tr>
<tr>
<td>2</td>
<td>CPS1</td>
<td>rs3770683 (intronic variant)</td>
<td>G</td>
<td>C</td>
<td>0.13</td>
<td>0.03</td>
<td>0.18</td>
<td>0.22</td>
<td>0.032</td>
<td>0.29</td>
<td>0.02</td>
<td>0.11</td>
<td>0.16</td>
</tr>
<tr>
<td>3</td>
<td>CYP2B6</td>
<td>rs2279343 (CYP2B6*4, *6, *7, *13, *16, *19, *20, *26, *34, *36, *37, *38; lys262Arg)</td>
<td>A</td>
<td>G</td>
<td>0.35</td>
<td>0.32</td>
<td>0.41</td>
<td>0.15</td>
<td>0.032</td>
<td>0.18 (Asian)</td>
<td>0.45 (YRI)</td>
<td>0.21 (CEU)</td>
<td>NA</td>
</tr>
<tr>
<td>4</td>
<td>CPS1</td>
<td>rs1800136 (Gln1382Gln)</td>
<td>G</td>
<td>A</td>
<td>0.35</td>
<td>0.06</td>
<td>0.17</td>
<td>0.33</td>
<td>0.011</td>
<td>0.01</td>
<td>0.25</td>
<td>0.25</td>
<td>0.23</td>
</tr>
<tr>
<td>5</td>
<td>ESR1</td>
<td>rs3770684 (intronic variant)</td>
<td>A</td>
<td>C</td>
<td>0.19</td>
<td>0.00</td>
<td>0.27</td>
<td>0.36</td>
<td>0.001</td>
<td>0.29</td>
<td>0.02</td>
<td>0.11</td>
<td>0.16</td>
</tr>
<tr>
<td>6</td>
<td>CYP2C19</td>
<td>rs2279343</td>
<td>C</td>
<td>T</td>
<td>0.35</td>
<td>0.56</td>
<td>0.44</td>
<td>0.28</td>
<td>0.031</td>
<td>0.31</td>
<td>0.17</td>
<td>0.15</td>
<td>0.36</td>
</tr>
<tr>
<td>7</td>
<td>HLA-B</td>
<td>rs1050462 (Val9Ile)</td>
<td>C</td>
<td>G</td>
<td>0.13</td>
<td>0.53</td>
<td>0.75</td>
<td>0.47</td>
<td>0.032</td>
<td>0.25</td>
<td>0.12</td>
<td>0.31</td>
<td>0.27</td>
</tr>
<tr>
<td>8</td>
<td>HLA-DQA1</td>
<td>rs1048087 (His94His)</td>
<td>C</td>
<td>T</td>
<td>0.45</td>
<td>0.32</td>
<td>0.22</td>
<td>0.53</td>
<td>0.011</td>
<td>NA</td>
<td>0.19 (YRI)</td>
<td>0.38 (CEU)</td>
<td>NA</td>
</tr>
<tr>
<td>9</td>
<td>HLA-DRB1</td>
<td>rs2308689 (Gly195Gly)</td>
<td>C</td>
<td>A</td>
<td>0.03</td>
<td>0.00</td>
<td>0.06</td>
<td>0.18</td>
<td>0.023</td>
<td>NA</td>
<td>0.50 (YRI)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>10</td>
<td>LDLR</td>
<td>rs2915966 (3’ UTR variant)</td>
<td>A</td>
<td>C</td>
<td>0.26</td>
<td>0.32</td>
<td>0.14</td>
<td>0.41</td>
<td>0.023</td>
<td>0.27</td>
<td>0.27</td>
<td>0.4</td>
<td>0.22</td>
</tr>
<tr>
<td>11</td>
<td>PGR</td>
<td>rs1171271</td>
<td>C</td>
<td>T</td>
<td>0.13</td>
<td>0.28</td>
<td>0.46</td>
<td>0.12</td>
<td>5.88E-07</td>
<td>0.08</td>
<td>0.28</td>
<td>0.16</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Differences in LEPR (not present in FDA panel) amongst Prakriti has been reported earlier. Exome sequencing has provided variation (rs1137101) information on this gene which is linked to drug response. This is included to highlight the merit of Ayurgenomics for identification of new genes to the panel of pharmacogenomics testing. The numbers in superscripts 1,2,3 indicate the significance in VK, PK, VP comparisons respectively.
<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Gene (class)</th>
<th>Drug substrate/activity</th>
<th>PGx relevance of gene</th>
<th>Consequence of variant wrt PGx relevance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CFTR</td>
<td>1) Anti-hyperglycemic- Glucose, 2) Cystic fibrosis drug- Ivacaftor, 3) Diuretic- Butamidine</td>
<td>CFTR protein is a chloride channel present at the surface of epithelial cells in multiple organs. 1) Anionic form of glyburide inhibits CFTR-CI channels. 2) Ivacaftor facilitates increased chloride transport by potentiating the channel-open probability of the G551D-CFTR protein. 3) Diuretic like bumetanide inhibit CFTR-Cl ion channels.</td>
<td>No pharmacogenomic significance of these variants have been reported. 1) Glibenclamide is an open channel blocker of CFTR. 2) Other polymorphism have been seen to affect the chloride ion channel functioning which leads to dysregulation of epithelial fluid transport (Clancy et al., 2014). 3) diuretics are open channel blockers of CFTR with distinct kinetics.</td>
</tr>
<tr>
<td>2</td>
<td>CPS1</td>
<td>Anti-epileptic- Valproic Acid</td>
<td>The CPS1 encodes a mitochondrial enzyme that catalyzes the conversion of ammonia to urea in the liver. CPS1 gene expression can be altered by epigenetic mechanisms. Therefore valproic acid may influence CPS1 transcription and thus ammonia metabolism.</td>
<td>Pharmacogenomics of these variants have been reviewed. Other variant has been associated with increased risk of hyperammonemia when receiving combined treatment of valproic acid and other antiepileptics in people with epilepsy also increasing the susceptibility to persistent pulmonary hypertension, and susceptibility to venoocclusive disease after bone marrow transplantation (<a href="http://www.genecards.org/cgi-bin/carddisp.pl?gene=CPS1">http://www.genecards.org/cgi-bin/carddisp.pl?gene=CPS1</a>)</td>
</tr>
<tr>
<td>3</td>
<td>CYP2B6</td>
<td>1) Anti-depressant/smoke cessation aid- bupropion 2) Anti-retroviral- efavirenz 3) Opioid dependence (S-methadone)</td>
<td>Hydroxybupropion contributes to the pharmacologic effects of bupropion for smoking cessation. And bupropion is metabolised into hydroxybupropion by CYP2B6. The major route of efavirenz metabolism to 8-hydroxyefavirenz is predominantly by CYP2B6. The secondary metabolite 8,14-dihydroxy-efavirenz is also mediated by CYP2B6. Methadone is metabolized primarily by CYP2B6 enzyme.</td>
<td>Loss-of-function variant CYP2B18 *2/*2 (poor metabolizer phenotype) is associated with decreased antipateptide response and reduced clopidogrel active metabolite formation when exposed to clopidogrel in healthy individuals. Low dose of clopidogrel should be administered.</td>
</tr>
<tr>
<td>4</td>
<td>CYP2C9</td>
<td>1) Anti-platelet clopidogrel, aspirin , labetalol (α/ß adrenergic antagonist) 2) Anti-convulsants/Anti-epileptic clobazam, mephenytoin 3) Anti- depressant 1) HLA-B induces the cytotoxic T-lymphocyte activation in patients treated with carbamazepine/phenytoin, thereby increasing the risk of SJS/TEN. HLA-B presents peptides to immune cells and is imp. in immune recognition of pathogens. Polymorphism in HLA-B increases risk of a hypersensitivity reaction with abacavir.</td>
<td>Hepatic bioactivation of clopidogrel. Aspirin induces the activity of CYP2C19. Also involved in pharmacodynamics of aspirin. Pharmacokinetic interaction with labetalol.</td>
<td>Wild type vs CYP2C19*2 allele shows differences in acid inhibition and pharmacodynamics. It is also associated with increased response to the drug. Low dose recommended (Hunfeld et al., 2010). Allele T is not associated with risk of musculoskeletal pain when exposed to anastrozole and letrozole in women with early stage breast cancer. A significant increased susceptibility to hepatitis B virus-related liver cirrhosis has also been observed in individuals with this SNP (Yan et al., 2011).</td>
</tr>
</tbody>
</table>
3) Allopurinol induces drug hypersensitivity on interaction with HLA-B proteins.

2) HLA-B*57:01-positive, abacavir is not recommended. 3) Allopurinol is contraindicated in individuals with the HLA-B*58:01-positive phenotype may require reduced doses of phenytoin.

8 HLA-DQA1 1) Anti-cancer (immunosuppressive antimetabolite)- Azathioprine, Mercaptopurine 2) multiple sclerosis therapy- Interferon beta-1a & 1b

1) Mercaptopurine interaction with HLA-DQA1 variation makes the individuals risk of pancreatitis 2) b-interferon (IFN-b) affects the expression of cell surface molecules and cytokine levels and can contribute significantly to alter the therapeutic responses, making them susceptible to MS.

9 HLA-DRB1 1) Anti-Cancer- Lapatinib 2) Antibiotic- Amoxicillin, Clavulanate 3) Anti-retroviral Abacavir 4) xanthine oxidase inhibitor (Uric acid lowering)- Allopurinol

1) HLA-DRB1 variants are predictor of increased risk of lapatinib-induced liver injury 2) variants in this gene increases the risk of liver injury in patients. 3) Abacavir interaction with HLA-DRB1 gene induces hypersensitivity 4) Allopurinol induces hypersensitivity

1) No Pharmacogenomic significance has been reported for this variant. Other HLA-DRB1 variant is a predictor of increased risk of lapatinib-induced liver injury. 3) Hypersensitivity to abacavir, positively associated with glatiramer acetate treatment response in Russian multiple sclerosis patients, risk allele in antiepileptic-drug induced Stevens-Johnson syndrome in Han Chinese. 4) Allopurinol-Induced Hypersensitivity in Hematologic Malignancy, severe cutaneous adverse reactions caused by allopurinol, and many other toxicity related response has been studied associated with this gene. No pharmacological effect of these polymorphisms have been reported.

10 LDLR Lipid lowering statin - Atorvastatin, Pravastatin, Rosuvastatin, Simvastatin

The drugs inhibits the LDLR gene to lower the plasma cholesterol level No pharmacogenomic significance of this variant have been reported. This has been seen to be in LD with rs2738464, a SNP associated with plasma lipid profiles indicative of higher cardiovascular risk (Muallem et al., 2007).

11 PGR Synthetic progestational hormone- Levonorgestrel, Progesterone, Mifepristone, Dydrogesterone, Drospirenone

The drugs are agonists of PGR gene. They bind to the PGR receptor to reduce the progesterone concentration required for gestation No pharmacogenomic significance of this variant have been reported.
time of examination. These need to be taken into account along with inter-observer variability during examination. For instance, a feature such as height being tall, medium or short has to be contextualized taking the background into account. Therefore, we evolved a two phase method of Prakriti evaluation. The first dealt with clinical methods of Ayurveda through screening and detailed phenotyping questionnaires and then measurement of phenotypic attributes through objective methods (Table 3). The consideration for selection and development of these objective parameters are detailed below.

4.1. Anatomical feature assessments through anthropometry

Prakriti assessment considers various anatomical features related to body structure and composition like size, shape symmetry, length & breadths as well bulk and quality of musculature of different individual body parts (Table 3). In Ayurveda, anthropometric measurements are described under Pramana Pariksha wherein the measurement of different body parts like heights, breadths, circumferences and lengths are described in a personalized manner in terms of his/her own finger (Swa Angula Pramana). These are described to vary with respect to Prakriti, gender, age and ethnic background and are also considered for assessing health and metabolic status (Sharma, 1981, 1999). However, visual inspection can lead to subjectivity in assessment. In order to provide objectivity to this aspect of Prakriti assessment more than 50 direct anthropometric parameters are measured. These also provide several indices and ratios of body that reflect the overall proportionate composition of an individual. Several anthropometric parameters and indices are contemporarily used to evaluate health and nutritional status, disease risk, and body composition changes that occur over the adult life (Huxley et al., 2010;
<table>
<thead>
<tr>
<th>Sl No</th>
<th>Prakriti related measurable phenotypes</th>
<th>Phenotypic features</th>
<th>Parameters (N)</th>
<th>Instruments used</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Size, shape and symmetry through Anthropometry</td>
<td>Overall body frame as well as individual body parts: head, face, shoulder, upper limb, lower limb, chest, waist, hip, palm and foot</td>
<td>Height, weight, span (3) Measurements of individual parts: • Lengths (20) • breadths (11) • Widths (5) • Girth &amp; circumference (8)</td>
<td>Anthropometric rod, Anthropometric Tape, Spreading &amp; Sliding calipers and Harpenden skin-fold caliper</td>
</tr>
<tr>
<td>2</td>
<td>Attributes of skin through measurements of biophysical properties</td>
<td>Skin color, texture, nature and types from 3 to 5 regions representing different levels of exposure to environment</td>
<td>• Sebum content of the skin • Melanin index (pigmentation)/erythema index (redness) • Skin pH • Skin elasticity/Firmness • Roughness and gloss of skin • Trans-epidermal water loss (TEWL) • Hydration of the stratum corneum.</td>
<td>Sebumeter Mexameter Sebuneter Mexameter Vapometer Skin-pH-Meter Cutometer Skin Gloss Meter Moisture meter SC Moisture meter D</td>
</tr>
<tr>
<td>3</td>
<td>Autonomic nervous system functions through Heart rate variability</td>
<td>Autonomic nervous system measures that correlate with metabolic and physical activity and response to stress: reflected in perspiration, bowel movements, psychosocial behavior such as anger, irritability etc</td>
<td>• Galvanic skin resistance (GSR) • Electrocardiogram (ECG) • Time domain: Mean RR, Mean HR, STDHR, SDNN, RMSSD, NN50, pNN50, RRt, TINN • Frequency domain: Very Low Frequency(VLF), Low Frequency(LF), High Frequency(HF) • Total Power Total Power, Sympathovagal ratio (LF/HF)</td>
<td>(Courage-Khazaka &amp; Delfin Instruments) MP150 Data Acquisition System: Three channels GSR, ECG, PPG (BIOPAC)</td>
</tr>
<tr>
<td>4</td>
<td>Lung Function Test through Spirometry</td>
<td>Vital capacity as a surrogate of physical strength</td>
<td>• Forced Vital capacity (FVC) • Forced expired volume in the first second (FEV1) • FEV1/FVC ratio • Average flow during the middle half of the FVC (FEF 25–75%)</td>
<td>Spirometer (NDD easyone)</td>
</tr>
<tr>
<td>5</td>
<td>Taste transduction through electrogustometry</td>
<td>Taste perception: taste sensitivity in six different regions of the tongue</td>
<td>Taste threshold in response to electric pulse ranging from –6 to 34 dB for a duration of 1.5 s</td>
<td>TR-06 RION Electrogustometer (Sensonics)</td>
</tr>
</tbody>
</table>
Kato et al., 2008; Onat et al., 2004; Sagun et al., 2014). In addition to providing objectivity to Prakriti assessment, this is also anticipated to provide normative values for different ethnic populations, gender and geography. Standard protocols are followed for measurements and three readings for each parameter are taken from the non-dominant side of the body (except biceps girth) by well-trained field staff/technicians under supervision of Ayurveda physicians.

4.2. Objective assessment of skin phenotypes

Prakriti types differ with respect to various skin features related to nature, texture, color and appearance which is incorporated in detailed Prakriti assessment questionnaire. To provide objectivity to these observations in Prakriti we are using various known biophysical skin parameters which map to different attributes in the skin as provided in Table 3. These parameters are used by different research groups not only to measure inter-individual variability in healthy across age, gender & body sites (Man et al., 2009) and environment (Cravello and Ferri, 2008) but also disease conditions where these biophysical parameters are altered e.g., eczema, psoriasis, atopic dermatitis, ichthyosis, DM type 1 & sleep disorder etc. (Blichmann and Serup, 1987; Gupta also disease conditions where these biophysical parameters are altered etc. (Antelmi et al., 2004; Choi et al., 2006; Stein et al., 1997; Teisala et al., 2014). Heart rate variability (HRV), beat-to-beat variation in either heart rate or the duration of the R–R interval – the heart period, has become an important risk assessment tool (Billman, 2011). A reduced HRV is shown to be associated with a poorer prognosis for a wide range of clinical conditions whereas, robust periodic changes in R–R interval are often a hallmark of health. It is observed that subtle functional change of health impairment or commencement of pathological changes is shown by HRV which is usually not detected by clinical assessments (Seifert et al., 2014). We are therefore using heart rate variability (HRV), as a window to autonomic functions among individuals of different Prakriti types to check for any correlation. If the HRV pattern differ amongst Prakriti groups, it may serve as a clinical marker of Prakriti enabling prediction of pre-disease states.

Most commonly used analytical methods of HRV are Frequency domain (FD) and Time domain (TD) analysis (Table 1). Standard protocols for data acquisition and quality check are being followed. Once the subject is relaxed as inferred from GSR signal, the HRV test period is started. The test consists of three phases, first 5 min in relax supine position followed by 3 min tilted position (at 60°) achieved within 15 s followed by re-supine relax position for 5 min. GSR and ECG recordings are taken for entire duration of 13 min. The analysis of the data is carried out using Kubios HRV software package.

4.5. Taste threshold assessment through electrogustometry

Ayurveda texts have described six types of basic tastes (sweet, sour, salty, pungent, bitter and astringent) with different Prakriti individuals having differences in taste preferences and suitability. The Prakriti questionnaire also includes questions related to liking and suitability for tastes and the associated memory. Sensitivity of gustatory system could be modulated by a number of short term and long term factors such as body mass, gender, age, local and systemic diseases, personal habits such as alcohol, smoking, drug dependence etc. (Zverev, 2004). Inter-individual variability in taste perception thresholds have been shown to be associated with food intake and further differential expression of taste receptor is relevant in many patho-physiological conditions (Shah et al., 2009; Shin et al., 2011). Electrogustometry is well established as a clinical tool for the estimation of taste detection thresholds (Stillman et al., 2003). We have used electrogustometry in order to see if taste thresholds vary among different Prakriti individuals. Electrogustometer (TR-06) is a device through which a stimulus is provided in the form of electric current at different thresholds (–6 to 34 db) on six different regions on tongue surface. The six regions were selected on the basis of distribution of papillae on tongue. The current threshold of taste perception at which the subject feels the stimulus is recorded.

The subjects have given specific instructions like proper cleaning of mouth in the morning & maintain nil by mouth orally conditions one hour prior to the test. Other confounding factors such as surgery, chronic cold, and tobacco chewing habits etc. are also recorded for enabling correct interpretation of data. Threshold of taste perception for each site is recorded twice with a gap of 5 min. In case the difference among the values is ±2 db, data is acceptable.
Ayurgenomics project overview: Flow diagram of the various aspects of clinical phenotyping and biological sample processing that have been undertaken in the five cohorts. Study cohorts have been identified from five geographical locations where through community engagements healthy individuals are recruited for Ayurgenomics studies. Health assessment is carried out prior to screening of Prakriti using a short questionnaire after taking consent from the subject for participation in the study. During screening, oral saliva is collected for DNA isolation. After a preliminary Prakriti analysis, a subset of predominant as well as mixed Prakriti individuals is recalled for detailed clinical phenotyping and sample collection. Besides detailed Prakriti evaluations, objective assessment of different anatomical parts and systems is carried out (Table 1). Peripheral blood, stool and urine samples are collected and processed for downstream biochemical, molecular, omics analysis. The leads are validated in model systems. All the activities are carried out after ethical clearance at sites and at CSIR-IGIB following ICMR guidelines for biomedical research.
5. **Prakriti stratification of population in diverse cohorts of genetically homogeneous populations**

A detailed protocol has been developed for phenotyping individuals based on Prakriti methods and development of a biological repository for downstream omics studies. Fig. 2 provides a project overview illustrating the work-flow of the Ayurgenomics activities that are being carried out.

5.1. **Population identification**

Indian populations are extremely diverse with respect to ethnicity, geography and linguistic lineages. The study of Indian Genome Variation Consortium project on populations of contrasting ethnicity from different linguistic and geographical background revealed four major genetic clusters (Consortium, 2008). The Prakriti study has been undertaken in five populations that map to two broad genetic clusters from diverse ethnic and geographical regions. This is being carried out to determine the relative prevalence of Prakriti types in these populations as well as identification of common clinical and molecular correlates of Prakriti. Within the geographical location, large endogamous populations have been identified.

5.2. **Preparatory phase**

This phase includes obtaining ethical clearance for the study at all the collaborative sites as well as centrally at CSIR-IGIB for the overall project as per ICMR biomedical research guidelines. The infrastructure for subject recruitment, clinical phenotyping and sample collection has been developed at all the field sites. Extensive training has been carried out for project personnel with the standardized protocols developed at CSIR-TRISUTRA for (a) subject recruitment, health assessment and obtaining informed consent (b) Prakriti evaluation (c) detailed assessment of objective parameters. Besides, onsite trainings at various field sites have also been imparted at regular intervals. A manual for Prakriti evaluation has been prepared to ensure uniform method for Prakriti assessment at different field sites.

5.3. **Community engagement and subject recruitment**

During this activity, community people for example health and social workers fluent in the local language of the concerned populations are actively involved in the study. This is to ensure maximal and authentic information and also to help the study volunteers understand the purpose of carrying out such a study. It is ensured that the individuals are mostly unrelated and both males and females are being collected in equal numbers in the age group of 19–40 years. Since the study is embarked on healthy population, a health assessment performa taking into consideration the methods described for same in Ayurveda has been made. For effective participation in the project, extensive community engagement efforts have been made involving local members. Various sites have evolved different methods for recruitment for instance creating awareness of health and wellness through lectures, free health check-up camps, and door-to-door household visits. Once the subject qualifies through the health assessment, he/she is recruited for the study following consent of the volunteer to participate in the study. This is carried out following all the ethical guidelines. During screening, oral saliva samples from the healthy volunteers is also collected for DNA sample repository. Nearly 20,000 individuals from the five cohorts are being screened. This comprise 10,000 volunteers from VADU and nearly 2000 from each of the four other sites. The study also includes large multi-generational families as well as elderly population from VADU who have been recruited to study heritability and effect of age on Prakriti phenotypes. The study populations represent two major genetic clusters from five different geographical locations that is population from Indo-European and Dravidian background from northern, southern, eastern and western part of India. One of the uniqueness of this study is that the subjects not only in each cohort are from a genetically homogeneous population, but are stratified into different Prakriti groups which include the extreme and most contrasting constitution types that represent only 10% of the population.

5.4. **Prakriti screening**

**Prakriti** assessment is being carried out in two phases. A screening phase which involves administration of a short screening questionnaire that contains representative questions from a detailed **Prakriti** questionnaire. A subset of subjects comprising of predominant Dosha as well as mixed Dosha Prakriti identified from screening are evaluated in detail using a more comprehensive questionnaire (Prasher et al., 2008). Cross-validation of Prakriti assessment between Ayurveda clinicians is achieved during detailed evaluation of the screened subjects. This also provides a background data for the observed frequency of Prakriti phenotypes in the study populations.

Manifestation of Prakriti phenotypes is also dependent on age and gender. Therefore, we carried out an analysis to assess the frequency of these phenotypes with respect to different age groups across both the genders on 6000 individuals from five cohorts (Fig. 3). Count of different options for each questions (phenotypes such as graying/falling for scalp hair, deep/shallow in sleep quality) were taken and chi square tests of these counts with respect to gender and age groups was performed. A significant p-value (cut off 0.05) was indicative of a skewing of a particular phenotype in a given age group. Non-significant values indicated equal distribution of options across age groups. For instance, a comparison of between age groups of 19–30 and 31–40 in both males and females revealed no significant difference with respect to certain parameters such as appetite regularity, food amount consumption, bowel habit tendency, perspiration amount, weather tolerance, speaking style etc. There were some parameters that significantly varied with respect to age across different geographical locations. Some of them were obvious parameters such as skin appearance, scalp hair changes, body bulk and musculature in females and scalp hair changes as well as body structure and composition in males. Since all the parameters are crucial to Prakriti assessment, the ones that vary, need to be normalized/ weighted differently with respect to age group and geography before attributing them to Prakriti. At the same time it would be worthwhile to evolve objective measures for substantiating the visual assessment of such phenotypes.

5.5. **Detailed phenotypic assessment using Prakriti methods and objective parameters**

A subset of subjects comprising of predominant Dosha identified from screening are evaluated in detail. Nearly 2000 individuals are being recruited for detailed phenotyping following an extensive protocol (Fig. 1). Large numbers of parameters that can individually provide objectivity to Prakriti assessment have been included in the study and detailed objective phenotyping is carried out in the subjects that are selected after screening (Table 1 and Fig. 2). This takes approximately 5–6 h and includes re-administration of consent form, detailed Prakriti questionnaire as well as measurement of anthropometric features, biophysical parameters of skin, heart rate variability (HRV), spirometry and electrogustometry (EGM) (Table 1). One day prior to phenotyping,
the subjects are provided specific instructions to be followed with respect to diet and personal hygiene. The subject is asked to relax, avoid food, beverages & drinks (tea, coffee) as well as smoking and alcohol consumption before detailed phenotyping and sample collection. This is done to minimize confounding observations during objective phenotyping as well as sample collection. Recruitment of individuals with transient illness or medications are rescheduled. All tests are conducted in prescribed ambient temperature (22–25 °C) and humidity. All the equipments are calibrated and maintained as per standard protocol and based on manufacturer’s recommendations.

6. Development of a biological sample repository

A bio-bank repository for storage of different biological samples that are being collected from study volunteers is being developed along with the data repository of phenome stratification. During Prakriti screening, saliva samples of all subjects are being collected. The peripheral blood, stool and urine are collected for the subjects who undergo detailed phenotyping. The saliva samples are processed for DNA isolation whereas peripheral blood is processed for isolation of DNA, RNA, serum and plasma as well as blood cells. Additionally, DNA is also isolated from stool and oral saliva for metagenomic analysis. Urine samples are dried and stored on filter papers for assessment of metabolites. Uniform standard protocols have been developed at CSIR-TRISUTRA and extensive training has been provided to all the lab personnel involved in sample collection, processing and storage at each site. Each of the field sites has been equipped with the molecular biology lab for processing of samples and extraction of DNA, RNA etc.

A centralised system of sample storage and retrieval mechanism has been established in CSIR-TRISUTRA. Samples collected and stored at sites are transferred to central repository at regular intervals. A uniform labeling mechanism has been evolved for the multiple samples types for each subject. To maintain confidentiality of the personal information linked to subjects, an additional round of anonymization of samples is done before distribution of data for analysis and publication (Kaye et al., 2009).

One of the objectives of the project is to create a repository of

Fig. 3. Phenotypic spectrum from Ayurgenomics cohorts for different parameters assessed during Prakriti screening: A Chi square test has been carried to infer differences in the distribution of options in each parameter during administration of short Prakriti screening questionnaire in (A) Males (B) Females, between age groups of 19–30 and 31–40. The p-values (cutoff threshold p < 0.05) of the tests have been plotted. Points below the black horizontal line represent the parameters that do not show any significant difference. Different plots refer to cohorts from diverse locations. This exercise has been carried out to assess uniformity as well as region specific differences across all cohorts. It highlights that some parameters need to be assessed keeping gender, age and geography in context during Prakriti evaluations as well as the need for evolving objective measures for assessment for few attributes.

EBV transformed B-cell immortalized cell lines from peripheral blood samples of a subset of predominant Prakriti subjects. A state of the art lymphoblastoid cell lines (LCL) facility has been established with trained personnel and standardized SOPs. This would provide genetic material for downstream validation studies.

7. Genomics and molecular analysis of individuals classified by their Prakriti types in healthy and diseased states

7.1. Biochemical profiling

Over 60 biochemical tests that are used in routine diagnostics for measurements of health of different systems are also being undertaken in Prakriti subjects. This include complete blood cell counts, hematocrit, platelet functions, liver, renal, cardiovascular and pancreatic functions, iron metabolism, lipid profiles, thyroid profiles, calcium metabolism, immunoglobulins, stress markers, vitamin D, folate, homocysteine levels, apo- lipoproteins, serum proteins, micronutrients etc. This would not only provide normative values in different ethnic populations and geographical regions across genders but also allow us to identify pre-clinical thresholds in populations stratified by Prakriti.

7.2. Dissecting phenome through OMICS

7.2.1. Establishing the genetic homogeneity of study populations

The current genomics platform technologies provide enormous opportunities to decipher the phenotype to genotype links at multiple layers from genetic to metabolic (Table 4). In TRISUTRA we are dissecting Prakriti at various omics levels which include: Genome-wide variations in the form of SNPs, CNVs, epigenome (DNA methylation marks); transcriptome (gene expression, transcript counting and differential expression); regulome (miRNAs and other small non-coding RNAs); metagenome (gut microbial community) and metabolome (Urine metabolite profiles). A wide range of array based and sequencing technologies are being employed to generate the multitude of data in a representative set of predominant Prakriti types as well as subjects of mixed Prakriti. Prior to embarking on such large scale omics studies, the genetic homogeneity of each cohort is established through analysis of ~3000 neutral markers that are also represented in the IGVdb (igvbrowser.igib.res.in). This is primarily to rule out issues of differences in Prakriti due to population stratification. The multi-omics data would be used to compare and identify molecular signatures of Prakriti across diverse populations. In a parallel effort, signatures of adaptation with respect to nearly 23 geo-climatic parameters have been identified in IGV and Ayurgenomics cohorts that would enable correlation of Prakriti differentiating genetic variations with geo-climatic adaptation. Validation of these signatures in health and disease cohorts might enable identification of early actionable points for interventions and also guide stratification of subjects that might have different therapeutic or disease management requirement.

7.3. Integrated data analysis and discovery of predictive markers

In addition to correlation of Prakriti with individual omics data, we are also carrying out integrated analysis of the multi-scale and multi-dimensional phenomic and genomic data (Cenik et al., 2015; Chen et al., 2012; Civelek and Lusis, 2014; Mitra et al., 2013). For example the parameters range from gross anatomical features to electrophysiologic tests of organ function, biochemical profiles of body fluids to the elucidation of genetic profiles using DNA and RNA. We shall test our hypothesis of existence of latent variables in our data which might correlate with Prakriti types in Ayurveda.

The presence of latent variables would be reflected in strong inter-correlations amongst the parameters. The breaking down of some correlations in an extreme-phenotype might be an indicator of predisposition to a particular disease. In the same manner, formation of strong correlations amongst parameters might drive the physiology towards more deterministic outcomes, the presence of which is usually associated with disease states. A particular advantage of this kind of approach is in its potential for pre-clinical detection of disease and discovery of markers that could drive development of better therapies. The markers could be phenotypic, genetic, physiological, biochemical or even a combination of these. The most informative of these markers can then be used as features to develop better classification criteria and even in the development intelligent classifiers using machine learning approaches.

Different statistical and machine learning based methods are being adopted/developed to carry out integrative analysis. The power of such integrative analysis has been shown in systems immunology for identification and predicting viral vaccine response and also adverse events (Kidd, 2016; Sobolev et al., 2016). The hope is if Prakriti can be correlated at the molecular level to such system-wide differences, phenotypic assessment through this method could become an affordable and non-invasive method that can be adopted in stratified medicine (Sethi et al., 2011).

7.4. Validation of leads in model system

The translation of genome-scale maps of genomic and epigenomic markers to clinically relevant information would require functional validation of these leads. Such studies are being carried out in model systems where it is possible to study the physiology towards more deterministic outcomes, the presence of which is usually associated with disease states. A particular advantage of this kind of approach is in its potential for pre-clinical detection of disease and discovery of markers that could drive development of better therapies. The markers could be phenotypic, genetic, physiological, biochemical or even a combination of these. The most informative of these markers can then be used as features to develop better classification criteria and even in the development intelligent classifiers using machine learning approaches.

Different statistical and machine learning based methods are being adopted/developed to carry out integrative analysis. The power of such integrative analysis has been shown in systems immunology for identification and predicting viral vaccine response and also adverse events (Kidd, 2016; Sobolev et al., 2016). The hope is if Prakriti can be correlated at the molecular level to such system-wide differences, phenotypic assessment through this method could become an affordable and non-invasive method that can be adopted in stratified medicine (Sethi et al., 2011).

7.5. Development of data-analytic platform for Ayurgenomics

As evident from above, an immense amount of data is being generated in different activities of the Ayurgenomics project (Fig. 4). The data is not only large in size but also multi-dimensional as well as multi-scale from various cohorts. This requires big data analytic platform technologies for information processing (Auffray et al., 2016; Mardis, 2016; Marx, 2013). The database houses information on Prakriti screening, detailed clinical phenotyping, sample processing, sample coding, physical storage, retrieval and large scale multi-omics data. All these data in their raw format as well as processed state are stored. The various computational aspects have been categorized into three broad kinds of activities (1) Development of methods and a relational database to gather and parse information that is being continuously generated at the field sites which includes clinical phenotyping data and bio-repository information (2) Data storage and server solutions for managing raw as well as processed data from various omics platforms (3) Pipelines for data analysis including user interface for integrative analysis. It is seamlessly connected with multiple layers of security, check points as well as user accesses. Ayurgenomics analysis platform is being developed on Apache Hadoop, Apache Hive and PostgreSQL technologies (O’Driscoll et al., 2013). It is envisaged to enable users to query and retrieve information in a fast and efficient manner, to analyze multi-omics data and also to enable researchers to unravel hidden relationship among various data points.

7.6. Data sharing and attribution policy

In the TRISUTRA Ayurgenomics consortium many research areas in collaboration with experts from diverse interdisciplinary fields have been undertaken. World over such interdisciplinary and data driven approaches have evolved certain practices related to data sharing and attribution efforts (Kaye et al., 2009). Appropriate credit sharing mechanisms have been evolved to ensure that the collaborators enhance their expertize and research in the area of developments of novel methods, algorithms, for joint or independent analysis. A set of guidelines have been mutually

![Data analytics platform for Ayurgenomics: A comprehensive overview of the diverse spectrum of activities towards which storage solutions, databases, analysis pipelines as well as user interfaces have been developed. A centralized repository for housing data related to clinical phenotyping, samples biobank, raw and processed Omics data as well as a user interface for analysis and visualization has been developed.](image)
Box 1–: TRISUTRA Ayurgenomics consortium

1. Project Conceptualization, Planning and Implementation: B Prashera*, M Mukerji*
2. Phenome Stratification: B Prashera

D. Phenomic data organization, QC and analysis: A Kumar*, BK Khuntia*, P Tiwari1, a
E. Predictive Modeling of Prakriti: P Tiwari1- a, R Kutum*, TP Sethi*, D Dash*, M Mukerji*; S Ghoshk

5. OMiCs experimental profiling (Exome, Genome, Epigenome, Transcriptome, Microbiome): M Mukerji*, B Prashera*, R Pandey*

A. Genotyping: B Varma*, A Tripathi*, R Thomas*, P Tiwari2, a, P Gautam*, R Kukreti*a
B. Gene Expression: S Aggarwal*, R Pandey*
D. Microbiome & Metagenomics: NS Chauhan*, J Kumar*, MK Verma*
e. Epigenomics: B Panda*, PGB Krishna*, AH Hariharan*, NM Krishnan*

8. Integrative Analysis: M Mukerji*, B Prasher*, D Dash*, R Kutum*, P Tiwari1, a, H Saxena*
9. Validation of leads in Model Systems: B Prasher*; M Mukerji*


evolved among all the participants. The roles and responsibilities, authorship and attribution in projects and publications have been addressed and clearly defined for all the ongoing projects (Box 1). The centralized repository of sample, data and protocols generated in consortium-wide endeavors is being housed and maintained in TRISUTRA with infinite access to all the collaborators for conducting the analysis towards fulfilling the objective of consortium. A prior consent or formal agreement with TRISUTRA would be required for any new scientific collaboration that involves the use of the resource of TRISUTRA. As the new dimensions of research would get added, members of TRISUTRA community are likely to increase and their names would be included as they become part of the TRISUTRA consortium. The leads from the Ayurgenomics data would be made publically available through scientific publications, popular articles and an Ayurgenomics browser. The discoveries arising out of the TRISUTRA research activities will be IPR protected and for their translational/wide spread use, partnership with relevant stakeholders would be undertaken.

8. Conclusion

It is envisaged that the leads from TRISUTRA would be useful in diverse aspects of health and disease management in common and complex disorders. It is anticipated that there would be several translational leads which would be attractive to health and nutrition industry. For example, identification of molecular correlates of Prakriti would be important in development of predictive and prognostic markers of disease as well as therapeutic response. The clinical markers could be important for development of formulations that target different Prakriti subtypes. Additionally, identification of molecular markers that could correlate with the outcome of Ayurveda therapeutic interventions including panchkarma, would increase the acceptability and outreach for such treatments. Research in this area could also complement drug target identification and preclinical trials. If we are successful in providing molecular evidence that different Prakriti have different baseline thresholds it would enable development of affordable health care solution through early screening in community health programs for early interventions.

Acknowledgment

The work was supported by grant (MLP901) from Council of Scientific and Industrial Research (CSIR) Govt of India for the project entitled “Setting up of a CSIR Unit- TRISUTRA (Translational Research and Innovative Science ThRough Ayurgenomics)”. Authors acknowledge contribution from all field staffs and medical/para medical/administrative staffs who worked painstakingly on the study. Authors acknowledge study population from KEMHRC-VADU, Pune, IPGTR&A, Gujarat Ayurved University, JB Roy State Ayurvedic MCH, Kolkata, Jamnagar, KLEU’s BMK Ayurveda Mahavidyalaya, Belgaum, CBPACS, Delhi areas for their participation in the study. TRISUTRA also acknowledges Prof. Samir Brahmachari and Dr Rajesh Gokhale CSIR-IGIB, for mentorship. TRISUTRA acknowledges Dr. Mohd Faruq for support in Next Generation Sequencing Facility. Authors acknowledge Dr Ram Niwas Prasher (MD Ayurveda) for extensive discussion on therapeutic interventions of Ayurveda and valuable suggestion. TRISUTRA also acknowledges CSIR-IGIB for administrative, infrastructure and IT support.

References

Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E.,


Varanasi.}


