THE IIOAB JOURNAL

VOLUME 1 : NO 2 : JULY 2010 : ISSN 0976-3104

Institute of Integrative Omics and Applied Biotechnology Journal

Dear Esteemed Readers, Authors, and Colleagues,

I hope this letter finds you in good health and high spirits. It is my distinct pleasure to address you as the Editor-in-Chief of Integrative Omics and Applied Biotechnology (IIOAB) Journal, a multidisciplinary scientific journal that has always placed a profound emphasis on nurturing the involvement of young scientists and championing the significance of an interdisciplinary approach.

At Integrative Omics and Applied Biotechnology (IIOAB) Journal, we firmly believe in the transformative power of science and innovation, and we recognize that it is the vigor and enthusiasm of young minds that often drive the most groundbreaking discoveries. We actively encourage students, early-career researchers, and scientists to submit their work and engage in meaningful discourse within the pages of our journal. We take pride in providing a platform for these emerging researchers to share their novel ideas and findings with the broader scientific community.

In today's rapidly evolving scientific landscape, it is increasingly evident that the challenges we face require a collaborative and interdisciplinary approach. The most complex problems demand a diverse set of perspectives and expertise. Integrative Omics and Applied Biotechnology (IIOAB) Journal has consistently promoted and celebrated this multidisciplinary ethos. We believe that by crossing traditional disciplinary boundaries, we can unlock new avenues for discovery, innovation, and progress. This philosophy has been at the heart of our journal's mission, and we remain dedicated to publishing research that exemplifies the power of interdisciplinary collaboration.

Our journal continues to serve as a hub for knowledge exchange, providing a platform for researchers from various fields to come together and share their insights, experiences, and research outcomes. The collaborative spirit within our community is truly inspiring, and I am immensely proud of the role that IIOAB journal plays in fostering such partnerships.

As we move forward, I encourage each and every one of you to continue supporting our mission. Whether you are a seasoned researcher, a young scientist embarking on your career, or a reader with a thirst for knowledge, your involvement in our journal is invaluable. By working together and embracing interdisciplinary perspectives, we can address the most pressing challenges facing humanity, from climate change and public health to technological advancements and social issues.

I would like to extend my gratitude to our authors, reviewers, editorial board members, and readers for their unwavering support. Your dedication is what makes IIOAB Journal the thriving scientific community it is today. Together, we will continue to explore the frontiers of knowledge and pioneer new approaches to solving the world's most complex problems.

Thank you for being a part of our journey, and for your commitment to advancing science through the pages of IIOAB Journal.

Yours sincerely,

Vasco Azevedo

Vasco Azevedo, Editor-in-Chief
Integrative Omics and Applied Biotechnology
(IIOAB) Journal

Prof. Vasco Azevedo
Federal University of Minas Gerais
Brazil

Editor-in-Chief

Integrative Omics and Applied Biotechnology (IIOAB) Journal Editorial Board:

The IIOAB Journal

ISSN: 0976-3104

RESEARCH: PHARMACOGENOMICS

GENETIC ADDICTION RISK SCORE (GARS) ANALYSIS: EXPLORATORY DEVELOPMENT OF POLYMORPHIC RISK ALLELES IN POLY-DRUG ADDICTED MALES

Kenneth Blum [1,2,5,6*], John Giordano[2], Siobhan Morse[2], Yiyun Liu[1], Jai Tan[3], Abdalla Bowirrat[4], Andrew Smolen[5], Roger Waite[6], William Downs[6], Margaret Madigan[6], Mallory Kerner[7], Frank Fornari[8], Eric Stice[9], Eric Braverman[10], David Miller[11], and John Bailey[1]

[1]*Department of Psychiatry, University of Florida, College of Medicine, Gainesville, Florida, USA*
[2]*Department of Holistic Medicine , G & G Holistic Addiction Treatment Center, North Miami Beach Florida, USA*
[3]*Medical Image Processing Group, Institute of Automation, Chinese Academy of Sciences, Beijing, CHINA & Life Sciences Research Center, School of Life Sciences and Technology, Xidian University, Xi'an, Shaanx, CHINA*
[4]*Clinical Neuroscience and Population Genetics, Ziv Government Medical Center, Safed, ISRAEL*
[5]*Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado, USA*
[6]*Department of Neuropsychiatric Genetics, Reward Deficiency Solutions, LLC., San Diego, California, USA*
[7]*Department of Clinical Neurology, Path Foundation, NY, New York, New York, USA*
[8]*Dominion Diagnostics, Kingstown,Rhode Island. USA*
[9]*Oregon Research Institute, Eugene, Oregon, USA*
[10]*Department of Neurological Surgery, Weil Cornell College of Medicine, New York, New York, USA*
[11]*LifeStream, Inc. Prescott, Arizona, USA*

Received on: 20th-June-2010; **Accepted on:** 8th-July-2010; **Published on:** 8th-July-2010.
[*]**Corresponding author:** Email: drd2gene@aol.com Tel: +1 619-890-2167; Fax: +1 619-236-3316

ABSTRACT

There is a need to classify patients at genetic risk for drug seeking behavior prior to or upon entry to residential and or non-residential chemical dependency programs. We have determined based on a literature review, that there are seven risk alleles associated with six candidate genes that were studied in this patient population of recovering poly-drug abusers. To determine risk severity of these 26 patients we calculated the percentage of prevalence of the risk alleles and provided a severity score based on percentage of these alleles. Subjects carry the following risk alleles: DRD2=A1; SLC6A3 (DAT) =10R; DRD4=3R or 7R; 5HTTIRP = L or LA; MAO= 3R; and COMT=G. As depicted in table 2 low severity (LS) = 1-36%; Moderate Severity =37-50%, and High severity = 51-100%. We studied two distinct ethnic populations group 1 consisted of 16 male Caucasian psycho stimulant addicts and group 2 consisted of 10 Chinese heroin addicted males. Based on this model the 16 subjects tested have at least one risk allele or 100%. Out of the 16 subjects we found 50% (8) HS; 31% (5) MS; and 19% LS (3 subjects). These scores are then converted to a fraction and then represented as a Genetic Addiction Risk Score (GARS) whereby we found the average GARS to be: 0.28 low severity, 0.44 moderate severity and 0.58 high severity respectively. Therefore, using this GARS we found that 81% of the patients were at moderate to high risk for addictive behavior. Of particular interest we found that 56% of the subjects carried the DRD2 A1 allele (9/16). Out of the 9 Chinese heroin addicts [one patient not genotyped] (group 2) we found 11% (1) HS; 56% (5) MS; and 33% LS (3 subjects). These scores are then converted to a fraction and then represented as GARS whereby we found the average GARS to be: 0.28 Low Severity; 0.43 moderate severity and 0.54 high severity respectively. Therefore, using GARS we found that 67% of the patients were at moderate to high risk for addictive behavior. Of particular interest we found that 56% of the subjects carried the DRD2 A1 allele (5/9) similar to group 1. Statistical analysis revealed that the groups did not differ in

ISSN: 0976-3104

terms of overall severity (67 vs. 81%) in these two distinct populations. Combining these two independent study populations reveal that subjects entering a residential treatment facility for poly-drug abuse carry at least one risk allele (100%). We found 74% of the combined 25 subjects (Caucasian and Chinese) had a moderate to high GARS. Confirmation of these exploratory results and development of mathematical predictive values of these risk alleles are necessary before any meaningful interpretation of these results are to be considered.

Keywords: *Genetic Addiction Risk Score (GARS); polymorphic genes; Neurotransmitters; Dopamine; Reward Deficiency Syndrome (RDS)*

[I] INTRODUCTION

Over half a century of dedicated and rigorous scientific research on the meso-limbic system provided insight into the addictive brain and the neurogenetic mechanisms involved in man's quest for happiness. In brief, the site of the brain where one experiences feelings of well being is the meso-limbic system. This part of the brain has been termed the "reward center". Chemical messages including serotonin, enkephalins, GABA and dopamine (DA), work in concert to provide a net release of DA at the nucleus accumbens (NAc), a region in the mesolimbic system. It is well known that genes control the synthesis, vesicular storage, metabolism, receptor formation and neurotransmitter catabolism. The polymorphic-versions of these genes have certain variations which could lead to an impairment of the neurochemical events involved in the neuronal release of DA. The cascade of these neuronal events has been termed "Brain Reward Cascade" [1] **[Figure-1]**. A breakdown of this cascade will ultimately lead to a dysregulation and dysfunction of DA. Since DA has been established as the "pleasure molecule" and the "anti-stress molecule," any reduction in function could lead to reward deficiency and resultant aberrant substance seeking behavior and a lack of wellness [2].

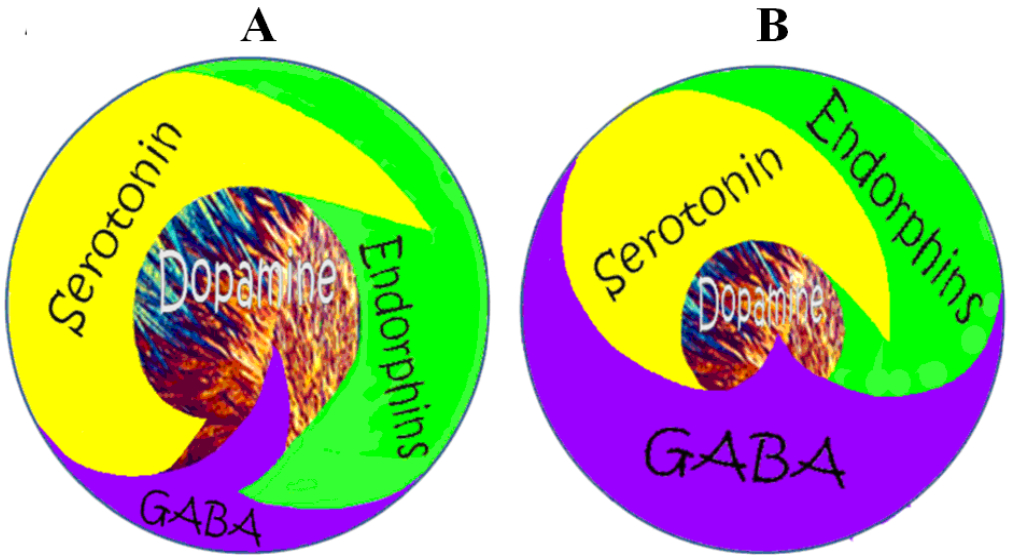


Fig: 1. Brain Reward Cascade. (A) Schematic represents the normal physiologic state of the neurotransmitter interaction at the mesolimbic region of the brain. Briefly in terms of the "Brain Reward Cascade" first coined by Blum and Kozlowski [90]: serotonin in the hypothalamus stimulates neuronal projections of methionine enkephalin in the hypothalamus which in turn inhibits the release of GABA in the substania nigra thereby allowing for the normal amount of Dopamine to be released at the NAc (reward site of Brain). **(B)** Represents hypodopaminergic function of the mesolimbic region of the brain. It is possible that the hypodopaminergic state is due to gene polymorphisms as well as environmental elements including both stress and neurotoxicity from aberrant abuse of psychoactive drugs (*i.e. alcohol, heroin, cocaine etc).* Genetic variables could include serotonergic genes (serotonergic receptors [5HT2a]; serotonin transporter 5HTlPR); endorphinergic genes (mu OPRM1 gene; proenkephalin (PENK) [PENK polymorphic 3' UTR dinucleotide (CA) repeats}; GABergic gene (GABRB3) and dopaminergic genes (ANKKI Taq A; DRD2 C957T, DRD4 7R, COMT Val/met substation, MAO-A uVNTR, and SLC6A3 9 or 10R). Any of these genetic and or environmental impairments could result in reduced release of dopamine and or reduced number of dopaminergic receptors.

ISSN: 0976-3104

Homo sapiens are biologically predisposed to drink, eat, reproduce and desire pleasurable experiences. Impairment in the mechanisms involved in these natural processes lead to multiple impulsive, compulsive and addictive behaviors governed by genetic polymorphic antecedents. While there are a plethora of genetic variations at the level of mesolimbic activity, polymorphisms of the serotonergic- 2A receptor (5-HTT2a); serotonergic transporter (5HTTLPR**);** (dopamine D2 receptor (DRD2), Dopamine D4 receptor (DRD4) ; Dopamine transporter (DAT1); and the Catechol-o-methyl –transferase (COMT) , monoamine –oxidase (MOA) genes as well as other candidate genes predispose individuals to excessive cravings and resultant aberrant behaviors [3].

An umbrella term to describe the common genetic antecedents of multiple impulsive, compulsive and addictive behaviors is Reward Deficiency Syndrome (RDS). Individuals possessing a paucity of serotonergic and/or dopaminergic receptors and an increased rate of synaptic DA catabolism, due to high catabolic genotype of the COMT gene, or high MOA activity are predisposed to self-medicating with any substance or behavior that will activate DA release including alcohol, opiates, psychostimulants, nicotine, glucose, gambling, sex, and even excessive internet gaming, among others [4]. Use of most drugs of abuse, including alcohol, is associated with release of dopamine in the mesocorticolimbic system or "reward pathway of the brain [5]. Activation of this dopaminergic system induces feelings of reward and pleasure [6, 7]. However, reduced activity of the dopamine system (hypodopaminergic functioning) can trigger drug-seeking behavior [8, 9]. Variant alleles can induce hypodopaminergic functioning through reduced dopamine receptor density, blunted response to dopamine, or enhanced dopamine catabolism in the reward pathway [10]. Possibly, cessation of chronic drug use induces a hypodopaminergic state that prompts drug-seeking behavior in an attempt to address the withdrawal –induced state [11].

Acute utilization of these substances can induce a feeling of well being. But, unfortunately sustained and prolonged abuse leads to a toxic pseudo feeling of well being resulting in tolerance and dis-ease or discomfort. Thus, low DA receptors due to carrying the DRD2 A1 allelic genotype results in excessive cravings and consequential behavior, whereas normal or high DA receptors results in low craving induced behavior. In terms of preventing substance abuse, or excessive glucose craving, one goal would be to induce a proliferation of DA D2 receptors in genetically prone individuals [12]. Experiments in vitro have shown that constant stimulation of the DA receptor system via a known D2 agonist in low doses results in significant proliferation of D2 receptors in spite of genetic antecedents [13]. In essence, D2 receptor stimulation signals negative feedback mechanisms in the mesolimbic system to induce mRNA expression causing proliferation of D2 receptors. This molecular finding serves as the basis to naturally induce DA release to also cause the same induction of D2-directed mRNA and thus proliferation of D2 receptors in the human. This proliferation of D2 receptors in turn, will induce the attenuation of craving behavior. In fact this has been proven with work showing DNA–directed over-expression (a form of gene therapy) of the DRD2 receptors and significant reduction in both alcohol and cocaine craving-induced behavior in animals [14, 15].

These observations are the basis for the development of a functional hypothesis of drug –seeking and drug use. The hypothesis is that the presence of a hypodopaminergic state, regardless of the source, is a primary cause of drug –seeking behavior. Thus, genetic polymorphisms that induce hypodopaminergic functioning may be the causal mechanism of a genetic predisposition to chronic drug use and relapse [12]. Finally, utilizing the long term dopaminergic activation approach will ultimately lead to a common safe and effective modality to treat RDS behaviors including Substance Use Disorders (SUD), Attention Deficit Hyperactivity Disorder (ADHD), and Obesity among other reward deficient aberrant behaviors.

Support for the impulsive nature of individuals possessing dopaminergic gene variants is derived from a number of important studies illustrating the genetic risk for drug-seeking behaviors based on association and linkage studies implicating these alleles as risk antecedents having impact in the mesocorticolimbic system [12].

1.1 D2 dopamine receptor gene (DRD2)

The dopamine D2 receptor gene (DRD2) first associated by Blum *et al* [17] with severe alcoholism is the most widely studied candidate gene in psychiatric genetics. The ***Taq***1 A is a single nucleotide polymorphism (SNP rs: 1800497) originally thought to be located at the 3' untranslated region of the DRD2 but now has been shown to be located within exon 8 of an adjacent gene, the ankyrin repeat and kinase domain containing 1 (ANKK1) [18]. Importantly, while there may be distict differences in function, Neville et al [18] suggest that the miss-location of the ***Taq***1 A may be attributable to the ANKKI and the DRD2 being on the same haplotype or the ANKKI being involved in reward processing through a signal transduction pathway. The ANKKI and the DRD2 gene polymorphisms may have distinct different actions with regard to brain function as has been noted in recent experiments and fear related conditioning in alcoholics [19, 20]. Grandy *et al.* [21] reported on the presence of the two alleles of the ***Taq***1 A: the A1 and A2 . Presence of the $A1^+$ genotype (A1/A1, A1 /A2) compared to the A^- genotype (A2/A2), is associated with reduced D2 receptor density [22, 23]. This reduction causes hypodopaminergic functioning in the dopamine reward pathway. Noble [24] in reviewing the literature concluded that research supports a predictive relationship from the $A1^+$ genotype to drug seeking behavior. This has been also discussed by Blum *et al* [3, 25] reporting that presence of the A+ genotype using **Bayesian** analysis has a predictive value of

IIOAB ISSN: 0976-3104

74% for a number of RDS behaviors. Other DRD2 polymorphisms such as the C [57T, a SNP (rs: 6277)] at exon 7 also associates with a number of RDS behaviors including drug use [26, 27, 28]. Compared to the T^- genotype (C/C), the T^+ genotype (T/T, T/C) is associated with reduced translation of DRD2 mRNA and diminished DRD2 mRNA [26], leading to reduced DRD2 density [27]. Hill *et al.* [28] has shown the predictive relationship between the T^+ allele and alcohol dependence. This results in hypodopaminergic function and is also a predictive risk allele.

The association of the DRD2 A1 allele in alcoholism is well established showing in a 10 year follow up that carriers of the DRD2 A1 allele have a higher rate of mortality compared to carriers of the A2 allele in alcohol dependent individuals [29]. There are 390 PUBMED reports [6/5/2010] providing significant support. The dopamine D2 receptor (DRD2) plays an important role in the reinforcing and motivating effects of ethanol. Several polymorphisms have been reported to effect receptor expression. The amount of DRD2, expressed in a given individual, is the result of the expression of both alleles, each representing a distinct haplotype.

Most recently, Kraschewski *et al*. [30] found that the haplotypes I-C-G-A2 and I-C-A-A1 occurred with a higher frequency in alcoholics [P=0.026, odds ratio (OR): 1.340; P=0.010, OR: 1.521, respectively]. The rare haplotype I-C-A-A2 occurred less often in alcoholics (P=0.010, OR: 0.507), and was also less often transmitted from parents to their affected offspring (1 vs.7). Among the subgroups, I-C-G-A2 and I-C-A-A1 had a higher frequency in Cloninger 1 alcoholics (P=0.083 and 0.001, OR: 1.917, respectively) and in alcoholics with a positive family history (P=0.031, OR: 1.478; P=0.073, respectively). Cloninger 2 alcoholics had a higher frequency of the rare haplotype D-T-A-A2 (P<0.001, OR: 4.614) always compared with controls. In patients with positive family history haplotype I-C-A-A2 (P=0.004, OR: 0.209), and in Cloninger 1 alcoholics haplotype I-T-A-A1 (P=0.045 OR: 0.460) were less often present. They confirmed the hypothesis that haplotypes, which are supposed to induce a low DRD2 expression, are associated with alcohol dependence. Furthermore, supposedly high-expressing haplotypes weakened or neutralized the action of low-expressing haplotypes.

1.2 D4 dopamine receptor gene (DRD4)

There is evidence that the length of the D4 dopamine receptor (DRD4) exon 3 variable number of tandem repeats (VNTR) affects DRD4 functioning by modulating the expression and efficiency of maturation of the receptor [31]. The 7 repeat (7R) VNTR requires significantly higher amounts of dopamine to produce a response of the same magnitude as other size VNTRs [32]. This reduced sensitivity or "dopamine resistance" leads to hypodopaminergic functioning. Thus 7R VNTR has been associated with substance –seeking behavior [32, 33]. However not all reports support this association [34]. Most recently Biederman *et al*. [35] evaluated a number of putative risk alleles using survival analysis, revealed that by 25 years of age 76% of subjects with a DRD4 7-repeat allele were estimated to have significantly more persistent ADHD compared with 66% of subjects without the risk allele. In contrast, there were no significant associations between the course of ADHD and the DAT1 10-repeat allele (P=0.94) and 5HTTLPR long allele. Their findings suggest that the DRD4 7-repeat allele is associated with a more persistent course of ADHD. This is consistent with our finding of the presence of the 7R DAT genotype in the heroin addict. Moreover in a study by Grzywacz *et al*. [36] which evaluated the role of dopamine D4 receptor (DRD4) exon 3 polymorphisms (48 bp VNTR) in the pathogenesis of alcoholism, they found significant differences in the short alleles (2-5 VNTR) frequencies between controls and patients with a history of delirium tremens and/or alcohol seizures (p = 0.043). A trend was also observed in the higher frequency of short alleles amongst individuals with an early age of onset of alcoholism (p = 0.063). The results of this study suggest that inherited short variants of DRD4 alleles (3R) may play a role in pathogenesis of alcohol dependence and carriers of the 4R may have a protective effect for alcoholism risk behaviors. It is of further interest that work from Kotler *et al*. [37] in heroin addicts illustrated that central dopaminergic pathways figure prominently in drug-mediated reinforcement including novelty seeking, suggesting that dopamine receptors are likely candidates for association with substance abuse in man. These researchers show that the 7-repeat allele is significantly over-represented in the opioid-dependent cohort and confers a relative risk of 2.46.

1.3 Dopamine Transporter gene (DAT1)

The dopamine transporter protein regulates dopamine –mediated neurotransmission by rapidly accumulating dopamine that has been released into the synapse [38]. The dopamine transporter gene (SLC6A3 or DAT1) is localized to chromosome 5p15.3. Moreover, within 3 non-coding region of DAT1 lies a VNTR polymorphism [38]. There are two important alleles that may independently increase risk for RDS behaviors. The 9 repeat (9R) VNTR has been shown to influence gene expression and to augment transcription of the dopamine transporter protein [39]. Therefore this results in an enhanced clearance of synaptic dopamine, yielding reduced levels of dopamine to activate postsynaptic neurons. Presence of the 9R VNTR has been linked to Substance Use Disorder (S.U.D.) [40] not consistently [41]. Moreover in terms of RDS behaviors, Cook *et al.* [42] was the first group that associated tandem repeats of the dopamine transporter gene (DAT) in the literature. While there have been some inconsistencies associated with the earlier results the evidence is mounting in favor of the view that the 10R allele of DAT is associated with high risk for ADHD in children and in adults alike. Specifically, Lee *et al.* [43] found consistent support in several studies, the non-additive association for the 10-repeat allele was significant for hyperactivity-impulsivity (HI) symptoms. However, consistent with other studies,

IIOAB ISSN: 0976-3104

exploratory analyses of the non-additive association of the 9-repeat allele of DAT1 with HI and oppositional defiant disorder (ODD) symptoms also were significant.

1.4 Catechol-O-methyltransferase (COMT)

The catechol-O-methyltransferase (COMT) is an enzyme involved in the metabolism of dopamine, adrenaline and noradrenaline. The Val158Met polymorphism of the COMT gene has been previously associated with a variability of the COMT activity, and alcoholism. Serý [44] found a relationship between the Val158Met polymorphism of the COMT gene and alcoholism in male subjects. Serý [44] found the significant difference between male alcoholics and male controls in allele and genotype frequencies ($p<0,007$; and $p<0$, 04 respectively. Interestingly in one of the subjects genotyped herein, who battles with heroin as an addiction while carrying the DRD2 A1 allele also carried the low enzyme COMT activity genotype (A/A). This is agreement with the work of Cao *et al.* [45] who did not find an association with the high G/G and heroin addiction. No differences in genotype and allele frequencies of 108 val/met polymorphism of COMT gene were observed between heroin-dependent subjects and normal controls (genotype-wise: chi-square=1.67, P=0.43; allele-wise: chi-square=1.23, P=0.27). No differences in genotype and allele frequencies of 900 Ins C/Del C polymorphism of COMT gene were observed between heroin-dependent subjects and normal controls (genotype-wise: chi-square=3.73, P=0.16; allele-wise: chi-square=0.76, P=0.38). While there is still some controversy regarding the COMT association with heroin addiction it was also interesting that the A allele of the val/met polymorphisms (-287 A/G) found by Cao *et al*. [45] was found to be much higher in heroin addicts than controls. Faster metabolism results in reduced dopamine availability at the synapse, which reduces postsynaptic activation, inducing hypodopaminergic functioning. Generally Vanderbergh et al [46] and Wang et al [47] support an association with the Val allele and SUD but others do not [48].

1.5 Monoamine -Oxidase A

Monoamine oxidase-A (MAOA) is a mitochondrial enzyme that degrades the neurotransmitters serotonin, norepinephrine, and dopamine. This system is involved with both psychological and physical functioning. The gene that encodes MAOA is found on the X chromosome and contains a polymorphism (MAOA-uVNTR) located 1.2 kb upstream of the MAOA coding sequences [49]. In this polymorphism, consisting of a 30-base pair repeated sequence, six allele variants containing either 2-, 3-, 3.5-, 4-, 5-, or 6-repeat copies have been identified [50]. Functional studies indicate that certain alleles may confer lower transcriptional efficiency than others. The 3-repeat variant conveys lower efficiency, whereas 3.5- and 4-repeat alleles result in higher efficiency [51]. The 3- and 4-repeat alleles are the most common, and to date there is less consensus regarding the transcriptional efficiency of the other less commonly occurring alleles (e.g., 2-, 5-, and 6-repeat). The primary role of MAOA in regulating monoamine turnover, and hence ultimately influencing levels of norepinephrine, dopamine, and serotonin, indicates that its gene is a highly plausible candidate for affecting individual differences in the manifestation of psychological traits and psychiatric disorders [52]. For example, recent evidence indicates that the MAOA gene may be associated with depression [53] and stress [54]. However, evidence regarding whether higher or lower MAOA gene transcriptional efficiency is positively associated with psychological pathology as been mixed. The low-activity 3-repeat allele of the MAOA-uVNTR polymorphism has been positively related to symptoms of antisocial personality [55] and cluster B personality disorders. Other studies, however, suggest that alleles associated with higher transcriptional efficiency are related to unhealthy psychological characteristics such as trait aggressiveness and impulsivity. Low MAO activity and the neurotransmitter dopamine are 2 important factors in the development of alcohol dependence. MAO is an important enzyme associated with the metabolism of biogenic amines. Therefore, Huang *et al.* [56] investigated whether the association between the dopamine D2 receptor (DRD2) gene and alcoholism is affected by different polymorphisms of the MAO type A (MAOA) gene. The genetic variant of the DRD2 gene was only associated with the anxiety, depression (ANX/DEP) ALC phenotype, and the genetic variant of the MAOA gene was associated with ALC. Subjects carrying the MAOA 3-repeat allele and genotype A1/A1 of the DRD2 were 3.48 times (95% confidence interval = 1.47-8.25) more likely to be ANX/DEP ALC than the subjects carrying the MAOA 3-repeat allele and DRD2 A2/A2 genotype. The MAOA gene may modify the association between the DRD2 gene and ANX/DEP ALC phenotype. Overall, Vanyukov et al. suggested that, although not definitive, variants in MAOA account for a small portion of the variance of SUD risk, possibly mediated by liability to early onset behavioral problems [57].

1.6 Serotonin Transporter gene

The human serotonin (5-hydroxytryptamine) transporter, encoded by the SLC6A4 gene on chromosome 17q11.1-q12, is the cellular reuptake site for serotonin and a site of action for several drugs with central nervous system effects, including both therapeutic agents (e.g. antidepressants) and drugs of abuse (e.g. cocaine). It is known that the serotonin transporter plays an important role in the metabolic cycle of a broad range of antidepressants, antipsychotics, anxiolytics, antiemetics, and anti-migraine drugs. Salz *et al.* [58] found an excess of -1438G and 5-HTTLPR L carriers in alcoholic patients in comparison to the heroin dependent group (OR (95% CI)=1.98 (1.13-3.45) and 1.92 (1.07-3.44), respectively). The A-1438G and 5-HTTLPR polymorphisms also interacted in distinguishing alcohol from heroin dependent patients (df) =10.21 (4), p=0.037). The association of -1438A/G with alcohol dependence was especially pronounced in the presence of 5-HTTLPR S/S, less evident with 5-HTTLPR L/S and not present with 5-HTTLPR

IIOAB ISSN: 0976-3104

L/L. SCL6A4 polymorphism haplotypes were similarly distributed in all three groups. Moreover, Seneviratne *et al.* [59] found that G allele carriers for rs1042173 were associated with significantly lower drinking intensity ($p = 0.0034$) compared to T-allele homozygotes. In HeLa cell cultures, the cells transfected with G allele showed a significantly higher mRNA and protein levels than the T allele-transfected cells. These findings suggest that the allelic variations of rs1042173 affect drinking intensity in alcoholics possibly by altering serotonin transporter expression levels. This provides additional support to the hypothesis that SLC6A4 polymorphisms play an important role in regulating propensity for severe drinking.

1.7 Combination of Genes and Addiction Risk

In general, inconsistencies in the literature involving association studies using single gene analysis prompted Conner *et al.* [60] and others to evaluate a number of dopaminergic gene polymorphisms as predictors of drug use in adolescents. We can't ignore the importance of neurochemical mechanisms involved in drug induced relapse behavior as suggested by Bossert *et al.* [61] understanding the interaction of multiple genes and environmental elements. These investigators have found using a drug relapse model, previously shown to induce relapse by re-exposing rats to heroin-associated contexts. After extinction of drug-reinforced responding in different contexts, re-exposure reinstates heroin seeking. This effect is attenuated by inhibition of glutamate transmission in the ventral tegmental area and medial accumbens shell, components of the mesolimbic dopamine system. This process enhances DA net release in the NAc. This fits well with Li's KARG addiction network map [62].

Since the initial finding of Blum *et al.* [17] showing positive association of a single gene DRD2 polymorphisms and severe alcoholism to date the replication, although favorable, has been fraught with inconsistent results. This has been true for other complex behaviors as well (NCI-NHGRI Working Group on Replication in Association studies *2007)*. Moreover, when gene-gene and environment interactions are tested the findings support the concept that complex gene –relationships may account for inconsistent findings across many different single gene studies [63].

There are many different reasons for inconsistencies in trying to predict drug use including single gene analysis, stratification of population, poor screened controls, gender–base differences, personality traits, co-morbidity of psychiatric disorders, positive and negative life events and even neurocognitive functioning [64, 65].

Thus, instead of continuing to evaluate single gene associations to predict future drug abuse, it occurred to us that we should embark on a study to evaluate multiple candidate gene candidates especially linked to the Brain Reward Cascade and hypodopaminergic functioning to gain a more complex but stronger predictive set of genetic antecedents. Our goal albeit exploratory in nature is to develop an informative panel to provide a means of stratifying or classifying patients entering a treatment facility as having low, moderate or high genetic predictive risk based on a number of known risk alleles. We are coining the term ***Genetic Addiction Risk Score*** (GARS) for purposes of study identification.

[II] MATERIALS AND METHODS

2.1 Subjects

The genotype data utilized in this paper is derived from previously published papers concerned with qEEG response from a natural Dopamine D2 agonist called Synaptose™ [64, 65] but the data set was never combined as accomplished herein.. The 16 patients were interviewed and evaluated for chemical dependence using a standard battery of diagnostic tests and questionnaires. The tests included the following: *Drug History Questionnaire; Physical Assessment, Urine Drug Tests; breathalyzer; Complete CBC blood test; and Symptom Severity Questionnaire*. The patients were determined to be substance dependent according to *Diagnostic and Statistical Manual* [DSM-IV] criteria. All patients were residential patients at G & G Holistic Addiction Treatment Center, North Miami Beach, Florida [14 patients] and the Bridging the Gaps, Winchester, Virginia [2 patients] treatment programs (30-90 day chemical dependence rehabilitation program). All subjects signed an approved consent form (approved by the IRB at PATH Foundation NY, New York, New York, registration # IRB00002334] and agreed to volunteer for this feasibility study. For protection of the patients the genotyping data conformed to standard HIPPA and GINA practices mandated by law.

Table-1 shows the demographics of the overall population including gender, race, age and length of abstinence. In this study there were a total of sixteen individuals. There were 16 Males and 0 females with a median age of 29.5 ± 8.8 SD years. The population breakdown was as follows: 87.5% Caucasian, and 12.5 Percent Hispanic. The average number of months abstinent for the entire population was 9.5 ±23.3. There were 3 pure cocaine only addicts; 4 cocaine crack addicts; 9 cocaine plus other drugs of abuse (alcohol, opiates and marijuana).

Table 1. Demographics of all Caucasian subjects combined

	Median ±st.dev.	(min, max)	N (total = 16)
Age	29.5 ±8.80	(19, 48)	16
Clean time (months)	9.5 ±23.33	(2, 101)	16
Race = Caucasian			14
Race = Hispanic			2
Sex = Male			16
Primary Substance = Cocaine only			3
Primary Substance = Crack cocaine			4
Primary Substance = Cocaine + Other			9

In **Table-2** we have also included genotype data from a fMRI study in China evaluating the effects of Synaptose™ in ten heroin addicted Chinese males. **Table-2** provides demographic information pertaining to this group. Diagnosis of heroin dependence was also determined in this group using DSM-IV criteria and other behavioral instruments. There were 10 Males and 0 females with a median age of 33 ± 7.6 SD years. The population breakdown was as follows: 100% Chinese. The average

IIOAB ISSN: 0976-3104

number of months abstinent for the entire population was 16 ± 7.9. There were 10 pure heroin only addicts.

2.2 Genotyping

A brief description of the genotyping methods for the polymorphisms to be assayed in this project follows. All methods are routinely performed in the Institute of Behavioral genetics (IBG), Boulder, Colorado laboratory. Each patient was also genotyped for the following gene polymorphisms: ***MAOA-VNTR*, 5HTTLPR, SLC6A3, DRD4, *ANKKI. DRD2 TaqIA (rs1800497)* and the *COMT val^{158}met SNP (rs4680).*** Genotypes were scored by two investigators independently.

The dopamine transporter (DAT1, locus symbol SLC6A3, which maps to 5p15.3, contains a 40 base-pair Variable Number Tandem Repeat (VNTR) element consisting of 3-11 copies in the 3' untranslated region (UTR) of the gene [66]. The assay [67] is a modification of the method of Vandenbergh *et al.* [66]. Primer sequences were: Forward- 5'-TGTGGTGTAGGGAACGGCCTGAG-3'; and Reverse- 5'-CTTCCTGGAGGTCACGCT CAAGG-3'.

Table 2. Demographics of all Chinese subjects combined*

	Median ±st.dev.	(min, max)	N (total = 10)
Age	33 ± 7.57	(20, 44)	10
Clean time (months)	16 ± 7.91	(1, 24)	10
Race = Chinese			10
Sex = Male			10
Primary Substance = Heroin only			10
Primary Substance = Heroin + other			0

*One sample was eliminated because of low amplification so that genotyping was not possible.

The dopamine D4 receptor (DRD4), which maps to 11p15.5, contains a 48 bp VNTR polymorphism in the third exon [68], which consists of 2-11 repeats. The assay [67] is a modification of the method of Lerman, *et al.* (1998) [69]. Primer sequences were: Forward- 5'-VIC -GCT CAT GCT GCT GCT CTA CTG GGC-3'; and Reverse-5'-CTG CGG GTC TGC GGT GGA GTC TGG-3'.

Monoamine Oxidase A upstream VNTR (MAOA-uVNTR): The MAOA gene, which maps to Xp11.3-11.4, contains a 30 bp VNTR in the 5' regulatory region of the gene which has been shown to affect expression [70]. A genotype by environment interaction has been reported for this polymorphism [71]. The MAOA-u VNTR assay is a modification [72] of a published method [70]. Primer sequences were: Forward- 5'-ACAGCCTGACCG-TGGAGAAG-3'; and Reverse- 5'-GAACGTGACGCTCCATTCGGA-3'.

Serotonin Transporter-Linked Polymorphic region (5HTTLPR): The serotonin transporter **(5HTT, Locus Symbol SLC6A4)**, which maps to 17q11.1-17q12, contains a **43 bp insertion/deletion** (ins/del) polymorphism in the 5' regulatory region of the gene [73]. Due to an error in sequencing this was originally thought to be a 44 bp deletion. The long variant (**L**) has approximately three times the basal activity of the short promoter (**S**) with the deletion [74]. Primer sequences were: Forward- 5'-6FAM - ATG CCA GCA CCT AAC CCC TAA TGT - 3'; Reverse- 5'- GGA CCG CAA GGT GGG CGG GA - 3'.

Hu *et al.* (2005) [75] reported that a SNP (rs25531, A/G) in the Long form of 5HTTLPR may have functional significance: The more common L_A allele is associated with the reported higher basal activity, whereas the less common L_G allele has transcriptional activity no greater than the S. The SNP rs25531 is assayed by incubating the full length PCR product with the restriction endonuclease MspI.

For all of the above VNTR and ins/del polymorphisms, PCR reactions contained approximately 20 ng of DNA, 10% DMSO, 1.8 mM MgCl2, 200 µM deoxynucleotides, with 7'-deaza-2'-deoxyGTP substituted for one half of the dGTP, 400 nM forward and reverse primers and 1 unit of AmpliTaq Gold® polymerase, in a total volume of 20 µl. Amplification was performed using touchdown PCR [76]. After amplification, an aliquot of PCR product was combined with loading buffer containing size standard (Genescan 1200 Liz) and analyzed with an ABI PRISM® 3130 Genetic Analyzer. Genotypes were scored by two investigators independently.

DRD2 TaqIA (rs1800497): The gene encoding the dopamine D2 receptor maps to 11q23, and contains a polymorphic TaqI restriction endonuclease site located within exon of the adjacent ANKKI gene which was originally thought tb located in the 3' untranslated region of the gene. The A1 allele has been reported to reduce the amount of receptor protein [77]. This SNP is done using a Taqman (5'Nuclease) assay [78]. Primer and probe sequences were: Forward primer- 5'-GTGCAGCTCACTCCATCCT-3'; Reverse primer- 5'-GCAACACAGCCATCCTCAAAG-3'; A1 Probe- 5'- VIC-CCTGCCTTGACCAGC-NFQMGB-3'; A2 Probe- 5'- FAM-CTGCCTCGACCAGC-NFQMGB-3'.

COMT val^{158}met SNP (rs4680): The gene encoding Catechol-O-methyltransferase (COMT) maps to 22q11.21, and codes for both the membrane-bound and soluble forms [79] of the enzyme that metabolizes dopamine to 3-methoxy-4-hydroxyphenylethylamine [80]. An A→G mutation results in a valine to methionine substitution at codons 158/108, respectively. This amino acid substitution has been associated with a four-fold reduction in enzymatic activity [80]. The COMT SNP is assayed with a Taqman [78] method. Primer and probe sequences were: Forward Primer- 5'-TCGAGATCAACCCCGACTGT-3'; Reverse Primer- 5'-AACGGG-TCAGGCATGCA-3'; Val Probe- 5'-FAM-CCTTGTCCTTCACGCCAGCGA- NFQMGB-3'; Met Probe- 5'-VIC-ACCTTGTCCTTCATGCCAGCGAAAT- NFQMGB-3'.

Details, including primer sequences and specific PCR conditions may be found in Anchordoquy *et al.* [67], Haberstick and Smolen [78] and Haberstick *et al.* [72].

2.3 Addiction Risk Score

In terms of genotyping data we have determined based on literature review that there are seven risk alleles involved in the six candidate genes studied in this patient population. To determine severity of the 25 patients studied (one Chinese subject was eliminated from the analysis due to poor PCR amplification) we calculated the percentage of prevalence of the risk alleles and provided a severity score based on percentage of risk alleles present. Subjects that carry the following alleles: DRD2=A1; SLC6A3 (DAT) =10R; DRD4=3R or 7R; 5HTTlRP = L or L_A; MAO= 3R; and COMT=G. As depicted in **Table- 2** Low Severity (LS) = 1-36%; Moderate Severity (MS) =37-50%, and High Severity (HS) = 51-100%.

[III] RESULTS

The resultant genotyping is illustrated in **Table-3** of this report and represents a total of 16 patients (group1) identified as not only addicts but the type of drug of choice.

IIOAB ISSN: 0976-3104

Table: 3. Group 1 Resultant genotyping data for each Caucasian patient.

Subject #	MAOA uVNTR	5HTTLPR	5HTTLPR	SLC6A3	DRD4	DRD2	COMT	Any risk allele	SEVERITY* ARS
1	3R	S/L	S/LG	9R/10R	4R/4R	A1/A2	G/G	POSITIVE	0.46–MS
2	3R	S/L	S/LA	10R/10R	4R/7R	A2/A2	G/G	POSITIVE	0.62 –HS
3	3R	L/L	LA /LG	9R/9R	3R/4R	A1/A2	A/G	POSITIVE	0.57-HS
4	4R	S/L	S/LA	10R/10R	3R/7R	A2/A2	G/G	POSITIVE	0.46-MS
5	4R	L/L	LA/LA	10R/10R	4R/7R	A2/A2	A/G	POSITIVE	0.62 –HS
6	3R	S/S	S/S	9R/10R	4R/7R	A2/A2	A/G	POSITIVE	0.30 –LS
7	4R	S/L	S/LG	10R/10R	4R/4R	A1/A1	A/A	POSITIVE	0.38 –MS
8	4R	S/L	S/LA	9R/10R	3R/4R	A2/A2	A/A	POSITIVE	0.23-LS
9	3R	L/L	LA//LA	9R/9R	4R/7R	A2/A2	A/G	POSITIVE	0.54-HS
10	4R	L/L	LA/LA	9R/10R	4R/4R	A2/A2	G/G	POSITIVE	0.54 –HS
11	3R	S/L	S/ LA	9R/10R	4R/4R	A1/A2	G/G	POSITIVE	0.54-HS
12	4R	L/L	LA/LA	9R/10R	4R/4R	A1/A2	A/G	POSITIVE	0.54-HS
13	4R	S/L	S/ LA	10R/10R	4R/4R	A1/A2	A/G	POSITIVE	0.46 –MS
14	4R	S/S	S/S	9R/10R	4R/4R	A1/A2	G/G	POSITIVE	0.30-LS
15	3R	L/L	LA / LA	10R/10R	4R/4R	A1/A2	A/G	POSITIVE	0.69 –HS
16	4R	S/L	S/LA	10R/10R	4R/7R	A1/A2	A/A	POSITIVE	0.46-MS

Severity percentage: LS =19; MS=31; HS= 50
Average GARS score: LS= 0.28; MS=0 .44; HS =0 .58
Prevalence of DRD2 A1 allele = 56%
Percentage of Moderate and High Severity= 81

In terms of genotyping data we have determined based on literature review that there are seven risk alleles involved in the six candidate genes studies in this patient population. To determine severity of the 16 patients studied we calculated the percentage of prevalence of the risk alleles and provided a severity score based on percentage of risk alleles present. Subjects that carry the following alleles: DRD2=A1; SLC6A3 (DAT)=10R; DRD4=3R or 7R; 5HTTlRP = L or L_A; MAO= 3R; and COMT=G. As depicted in **Table-2** low severity (LS) = 1-36%; Moderate Severity MS) =37-50%, and High Severity (HS) = 51-100%. Based on this model 16 subjects tested have at least one risk allele or 100%. Out of the 16 subjects we found 50% (8) HS; 31% (5) MS; and 19% LS (3 subjects). These scores are then converted to a fraction and then represented as an GARS whereby we found the average GARS to be: 0.28 Low Severity; 0.44. moderate severity and 0.58 high severity respectively. Therefore, using GARS we found that 81% of the patients were at moderate to high risk for addictive behavior. Of particular interest we found that 56% of the subjects carried the DRD2 A1 allele (9/16) **[Table-3]**.

Table 4. Group 2 Resultant genotyping data for each Chinese patient.

Subject #	MAOA uVNTR	5HTTLPR	5HTTLPR	SLC6A3	DRD4	DRD2	COMT	Any risk allele	SEVERITY* ARS
1	4R	S/L	S/L_A	10R/10R	4R/4R	A2/A2	A/A	POSITIVE	0.30–LS
2	3R	S/S	S/S	10R/10R	2R/4R	A1/A2	GAG	POSITIVE	0.38–MS
3	4R	S/S	S/S	10R/10R	3R/4R	A1/A2	G/G	POSITIVE	0.46-MS
4	3R	S/S	S/S	10R/10R	4R/6R	A2/A2	G/G	POSITIVE	0.38-MS
5	4R	S/S	S/S	10R/10R	4R/4R	A1/A2	A/G	POSITIVE	0.30–LS
6	3R	L/L	S/ L_G	10R/10R	4R/4R	A1/A2	ND	POSITIVE	0.45 –MS
7	4R	L/L	L_A/L_G	10R/10R	4R/4R	A1/A2	A/G	POSITIVE	0.54 –HS
8	4R	S/S	S/S	10R/10R	4R/5R	A2/A2	A/G	POSITIVE	0.23-LS
9	3R	S/L	S/L_A	10R/10R	2R/4R	A2/A2	A/G	POSITIVE	0.46-MS

Severity percentage: LS =33; MS=56; HS= 11
Average GARS score: LS= 0.28; MS=0 .43; HS =0 .54
Prevalence of DRD2 A1 allele = 56%
Percentage of Moderate and High Severity= 67

Moreover, data obtained from an on-going fMRI study in China (YL and JT) in nine heroin addicted males [see demographic **Table-2**] show similar genotype data **[Table-4]**. Based on this model 9 subjects tested (Group 2) have at least one risk allele or 100%. Out of the 9 subjects we found 11% (1) HS; 56% (5) MS; and 33% LS (3 subjects). These scores are then converted

IIOAB ISSN: 0976-3104

to a fraction and then represented as an GARS whereby we found the average GARS to be: 0.28 Low Severity; 0.43 moderate severity and 0.054 high severity respectively. Therefore, using GARS we found that 67% of the patients were at moderate to high risk for addictive behavior. Of particular interest we found that 56% of the subjects carried the DRD2 A1 allele (5/9) **[Table-4]**. Statistical analysis revealed that the groups did not differ in terms of overall severity (67 vs. 81%) in these two distinct populations. Using the z-test of proportions, the resulting z=0.79 with p=0.432. However a sample size of 228 for Group 1 and 128 for Group 2 to detect a significant difference between two populations with 81% and 67% risk by z-test at the 0.05 level with power of 80%.

Nevertheless, combining these two independent study populations (Group 1 and Group2) reveal that subjects entering a residential treatment facility for poly-drug abuse carry at least one risk allele (100%). We found that 74% of this combined 25 subjects (Caucasian and Chinese) had a moderate to high GARS.

[IV] DISCUSSION

While this exploratory study did not carry out any specific statistical analysis such as Baysian Theorem, Structural Equation Modeling or Recursive Partitioning (PR) the subject of work in progress the study was still informative. In terms of genotyping data we have determined that when multiple candidate genes are analyzed such as serotonergic- 2A receptor (5-HTT2a); serotonergic transportor (5HTTLPR); (dopamine D2 receptor (DRD2), Dopamine D4 receptor (DRD4); Dopamine transporter (DAT1); Catechol-o-methyl –transferase (COMT), and monoamine –oxidase (MOA) genes we found that 100% of all subjects carried at least one risk allele. Moreover this is the first time that anyone attempted to stratify or classify addiction risk by incorporating an algorithm formulation of combining a number of risk alleles by pre-assigning an allele as an risk allele having predictive value for drug use. For example it has been published earlier that the DRD2 A1 allele had a predictive value for all Reward Deficiency Syndrome (RDS) behaviors using Baysian statistics to have a high predictive value of 74.4%. [3] and reviewed by Bowirrat *et al.* [81]. It is of further interest that the subjects studied in this investigation had multiple drug abuse relapses and presented to in-patient residential treatment programs. Our preliminary finding of approximately 75% of these individuals having moderate to high GARS whereby only 25% had low GARS suggest a potential utility for pre-screening patients prior to a one-size fits all treatment plan. Clinically this may have real importance in understanding expectations of future success and the need for intensive treatment involving genomic solutions coupled with bio-holistic medical therapies [82].

The present exploratory study supported the hypothesis suggested earlier by us and others [60, 83] by identifying hypodopaminergic genotypes as the best predictor of drug abuse behavior in an adult and even more so in an adolescent population. This work is in agreement with Melis *et al.* [11] that identified a hypodopaminergic state as a causal mechanism in the development of SUD. This is consistent with a number of functioning Magnetic Resonance Imaging (fMRI) studies showing the importance of DRD2 levels by genotyping indicating that hypodopaminergic A1 genotype leads to blunted response and as such could lead to aberrant drug and or food seeking behavior [84, 85] while hyperdopaminergic A2 genotype serves as a protective factor against the development of drug disorders [86].

A further strength of this study is that we only used male subjects. de Courten –Myers *et al.* [87] have pointed out that one of the difficulties in replicating single gene associations with drug use disorder is sex –based or gender differences in neuro -chemistry and neuroanatomy. Moreover, Conner .et al. [60] suggested that males with hypodopaminergic functioning are more likely to abuse drugs that stimulate the mesocorticallmbic system than those with normal dopaminergic functioning. In contrast, females living in a negative environment are at increased risk (possibly not due to their genotypes) for using more drugs and even more types of drug which increase their risk for SUD.

Another strength of this exploratory study is that it is in agreement with the work of Conner *et al.* [60] confirming the importance of the cumulative effect of multiple genotypes coding for hypodopaminergic functioning, regardless of their genomic location, as a predictive method of drug use in males. Moreover, it extends the current literature, by suggesting for the first time a simple method using genetic testing to classify risk behavior in male patients seeking in-patient residential treatment.

The limitations of this study must be considered before interpreting the findings. This was only an exploratory study and as such a small sample size was utilized to obtain very preliminary data. This study showing positive association of a number of hypodopaminergic gene polymorphisms with drug abusing adults requires replication in a much larger population in both in-patient and out-patient facilities. The confirmatory studies must include both males and females. The studies should extend the population base to specific drugs of choice, ethnic groups, age and other risk taking behaviors. Certainly the frequency of drug seeking behavior must also be considered in future experiments. Using a SUD scale [88] may also improve the generality of these findings. Most importantly many more candidate genes should be included in the GARS panel. Blum *et al.* [89] has reported on a so called "Happiness Gene Map which includes a total of 30 genes. These genes influence how reward is interpreted in the brain .Another impotent caveat is that the expression of these gene polymorphisms may be significantly impacted by epigenetic effects due to environmental elements.

While it is understood that future work will analyze the best predictive candidate genes to secure a predictive GARS panel

IIOAB ISSN: 0976-3104

of genes utilizing a number of statistical tools such as recursive partitioning and Baysian predictive modeling techniques the need for such a genetic test in the Chemical Dependence field seem parsimonious. A major limitation herein is that larger sample size and the definitive association of these risk alleles with validated severity scales (i.e. treatment response, failures and number of years addicted) are warranted. There are at least three practical reasons for such a diagnostic test: 1) identifying those at risk prior to the onset of SUD providing early intervention and prevention of the negative outcomes from such use: 2) removal of denial and guilt and 3) genotype results could suggest different at risk individuals and programs could be tailored to a patients risk profile.

It is important to note that the severity of risk in the Caucasian seemed to be somewhat different when we only look at the percentage of high GARS. Specifically, 50% of the psychostimulant drug of choice dependent individuals (Caucasian) had a high GARS whereas only 11% of the Heroin addicted males (Chinese) had a high GARS. We do not have a reasonable explanation for this difference. However when both moderate and High GARS are combined for both groups we find that a total of 74% of these poly-drug abusers have a moderate to high GARS.

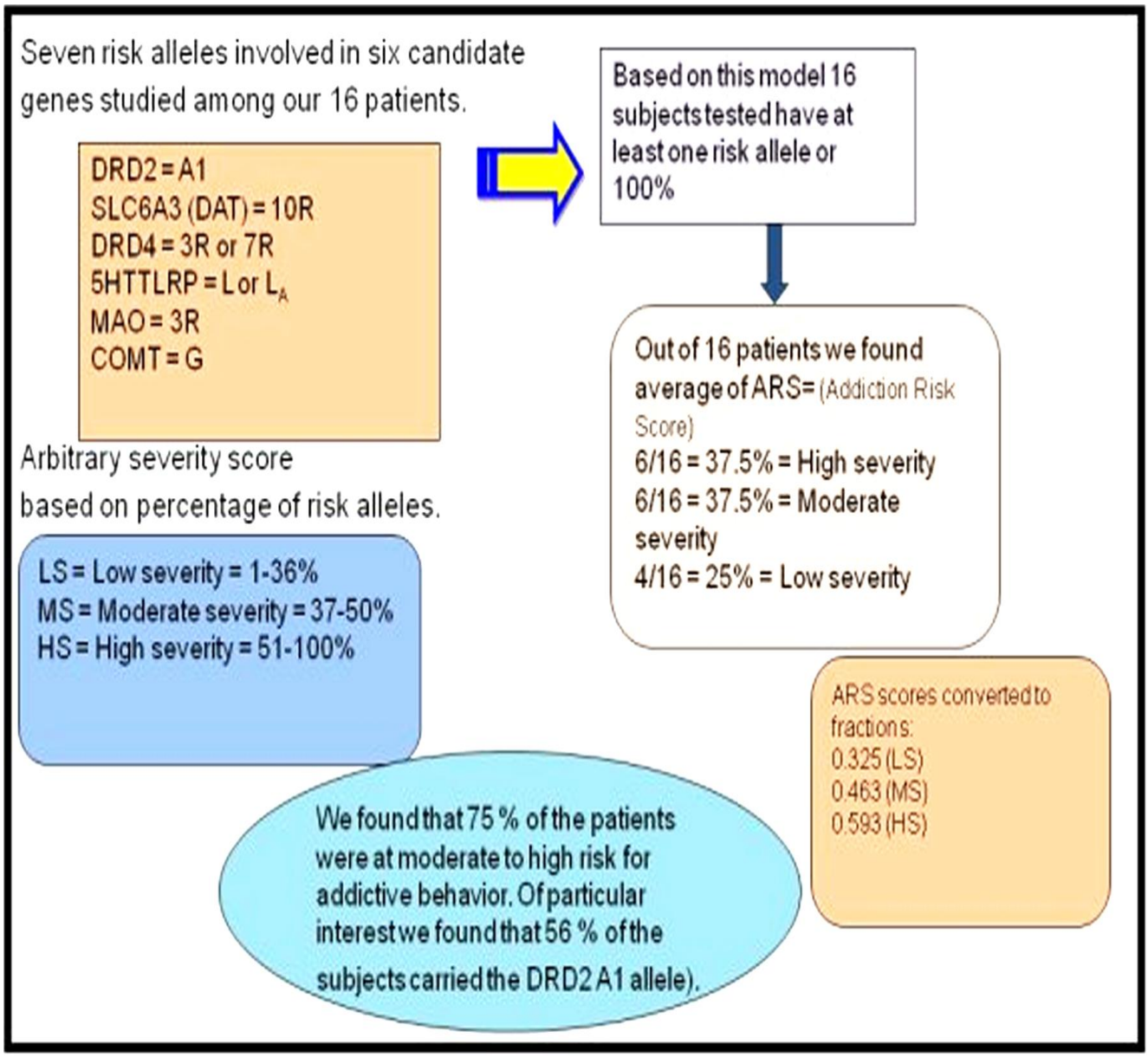


Fig: 2. Genetic Addiction Risk Score (GARS) Analysis: Exploratory Development of polymorphic risk alleles among 16 addicted patients. The figure does not display the results obtained for the Chinese samples.

IIOAB ISSN: 0976-3104

[V] CONCLUSION

The need to genetically test individuals especially at entry into a residential or even non-residential chemical dependency program has been suggested by scientists and clinicians alike here and abroad. In fact the most recent work of Conner *et al.* [60] has suggested the importance of multiple hypodopaminergic gene polymorphisms as a possible predictive tool to identify children at risk for problematic drug use prior to the onset of drug dependence. Our current exploratory study of only 16 Caucasians [as summarized in **Figure-2**] is in agreement with this prediction in terms of the development of a novel genetic test using an algorithm to determine the proposed GARS. To reiterate we found a high percentage (75%) of subjects carry a moderate to high GARS whereby 100% of individuals tested posses at least one risk allele tested. It is of some interest that in the Chinese population Group 2 only we found rare DRD4 alleles in this population such as 2R, 5R and 6R.

We are proposing, it is possible that the hypodopaminergic state is due to gene polymorphisms as well as environmental elements including both stress and neurotoxicity from aberrant abuse of psychoactive drugs (i.e .alcohol, heroin, cocaine etc). Genetic variables could include serotonergic genes (serotonergic receptors [5HT2a]; serotonin transporter 5HTlPR); endorphinergic genes (mu OPRM1 gene; proenkephalin (PENK) [PENK polymorphic 3' UTR dinucleotide (CA) repeats}; GABergic gene (GABRB3) and dopaminergic genes (ANKKI Taq A; DRD2 C957T, DRD4 7R, COMT Val/met substation, MAO-A uVNTR, and SLC3 9 or 10R). Any of these genetic and or environmental impairments could result in reduced release of dopamine and or reduced number of dopaminergic receptors.

We are proposing that following needed confirmation positive outcome of GARS will have prevention and treatment benefits in those probands afflicted with genetic antecedents to RDS seeking behaviors.

FINANCIAL DISCLOSURE

The following authors have financial conflicts based on patented technology related to the Genetic Addiction Risk Score (GARS) gene panel which has been licensed by Kenneth Blum to LifeGen, Inc. San Diego, California: Kenneth Blum, Roger L. Waite, B William Downs, Margaret Madigan, Abdalla Bowirrat, and David Miller.

ACKNOWLEDGEMENT

The authors are appreciative of the work and comments provided by the PATH Foundation NY staff. The authors are grateful to Stan Stokes of Bridging the Gaps and Mathew Manka, Debra Manka, and Merlene Miller of LifeStream Inc. Prescott, Arizona for permission to incorporate the gene data of two residential patients.

RESOURSE LINKS

The following Web site links are suggested for additional information: http:// www.addictionseearch.com; http://www.drugstrategies.org and http://www.rdsyndrome.com

REFERENCES

[1] Blum K, Cull JG, Braverman ER, Comings DE. [1996] Reward Deficiency Syndrome. *The American Scientist* 84: 132-145.

[2] Blum K, Braverman ER. [2000] Reward Deficiency Syndrome: A Biogenetic Model For the Diagnosis And Treatment Of Impulsive, Addictive and Compulsive Behaviors. *Journal Of Psychoactive Drugs* 32: 1-112 (Supplement).

[3] Blum K, Sheridan PJ, Wood RC, Braverman ER, et al. [1996] The D2 dopamine receptor gene as a determinant of reward deficiency syndrome. *J R Soc Med* 89: 396-400.

[4] Comings DE, Blum K. [2000] Reward deficiency syndrome: genetic aspects of behavioral disorders. *Prog Brain Res* 126: 325-41.

[5] Di Chiara G, Impereto A. [1988] Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic systems of freely moving rats. *Proc Natl Acad Sci U S A* 84: 1413-1416.

[6] Volkow ND, Fowler JS, Wang GJ. [2002] Role of dopamine in drug reinforcement and addiction in humans: results from imaging studies. *Behav Pharmacol* 13: 355-366.

[7] Eisenberg DT, Campbell B, Mackillop J, Lum JK, Wilson DS. [2007] Season of birth and dopamine receptor gene associations with impulsivity, sensation seeking and reproductive behaviors. *PLoS ONE* 2:e1216

[8] Volkow ND, Chang L, Wang GJ, et al. [2001] Low level of brain dopamine D2 receptors in methamphetamine abusers: Association with metabolism in the orbitofrontal cortex. *Am J Psychiat* 158: 377-382.

[9] Dackis CA, Gold MS, Davies RK, Sweeney DR. [1985]. Bromocriptine treatment for cocaine abuse: the dopamine depletion hypothesis. *Int J Psychiatry Med* 15 (2):125-135.

[10] Hietata J, West C, Syvalahti E, et al. [1994] Striatal D2 dopamine receptor binding characteristics in vivo in patients with alcohol dependence. *Psychopharmacol* 116:285-290.

[11] Melis M, Spigra S, & Diana M. [2005] The dopamine hypothesis of drug addiction: Hypodopaminergic State. *International Review of Neurobiology* 63: 101-154.

[12] Rothman RB, Blough BE, Baumann MH. [2007] Dopamine/serotonin releasers as potential medications for stimulant and alcohol addictions. *AAPS J* 9: E1-10.

[13] Boundy VA, Pacheco MA, Guan W, Molinoff PB. [1995] Agonists and antagonists differentially regulate the high affinity state of the D2L receptor in human embryonic kidney 293 cells. *Mol Pharmacol* 48: 956-964.

[14] Thanos PK, Michaelides M, Umegaki H, Volkow ND. [2008] D2R DNA transfer into the nucleus accumbens attenuates cocaine self-administration in rats. *Synapse* 62: 481-486.

[15] Thanos PK, Volkow ND, Freimuth P, Umegaki H, Ikari H, Roth G, Ingram DK, Hitzemann R. [2001] Overexpression

IIOAB ISSN: 0976-3104

of dopamine D2 receptors reduces alcohol self-administration. *J Neurochem* 78: 1094-1103.

[16] Merlo LJ, Gold MS. [2008] Special Report- Frontiers in psychiatric research: Addiction research: The state of the art in 2008. *Psychiatric Times* 25: 52-57.

[17] Blum K, Noble EP, Sheridan PJ, et al. [1990] Allelic association of human dopamine D2 receptor gene in alcoholism. *Journal of the American Medical Association* 263: 2055-2060.

[18] Neville MJ, Johnstone EC, Walton RT. [2004] Identification and characterization of ANKK1: a novel kinase gene closely linked to DRD2 on chromosome band 11q23.1. *Hum Mutat* 23: 540-545.

[19] Dick DM, Wang JC, Plunkett J, et al. [2007] Family-based association analyses of alcohol dependence phenotypes across DRD2 and neighboring gene ANKK1. *Alcohol Clin Exp Res* 31: 1645-1653.

[20] Huertas E, Ponce G, Koeneke MA, Poch C, España-Serrano L, Palomo T, Jiménez-Arriero MA, Hoenicka J. [2010] The D2 dopamine receptor gene variant C957T affects human fear conditioning and aversive priming. *Genes Brain Behav* 1: 103-109.

[21] Grandy DK, Litt M, Allen L, Bunzow JR, Marchionni M, Makam H, Reed L, Magenis RE, Civelli O.[1989] The human dopamine D2 receptor gene is located on chromosome 11 at q22-q23 and identifies a TaqI RFLP. *Am J Hum Genet* 45(5): 778-85.

[22] Noble EP, Blum K, Ritchie T, Montgomery A, Sheridan P. [1991] Allelic association of the D2 dopamine receptor gene with receptor-binding characteristics. *Archives of General Psychiatry* 48: 648-654.

[23] Pohjalainen T, Rinne JO, Någren K, Lehikoinen P, Anttila K, Syvälahti EK, Hietala J. [1998] The A1 allele of the human D2 dopamine receptor gene predicts low D2 receptor availability in healthy volunteers. *Mol Psychiatry* 3(3): 256-260.

[24] Noble EP. [2003] D2 Dopamine receptor gene in psychiatric and neurologic disorders and its phenotypes. *Am J Med Genet* 116B:103-125.

[25] Blum K, Braverman ER, Wood RC, Gill J, Li C, Chen TJ, Taub M, Montgomery AR, Sheridan PJ, Cull JG. [1996] Increased prevalence of the Taq I A1 allele of the dopamine receptor gene (DRD2) in obesity with comorbid substance use disorder: a preliminary report. *Pharmacogenetics* 6(4): 297-305.

[26] Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J, Gejman PV.[2003] Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum Mol Genet* 12: 205-216.

[27] Hirvonen M, Laakso A, Någren K, Rinne JO, Pohjalainen T, Hietala J. [2004] C957T polymorphism of the dopamine D2 receptor (DRD2) gene affects striatal DRD2 availability in vivo. *Mol Psychiatry* 91060-1.

[28] Hill SY, Hoffman EK, Zezza N, Thalamuthu A, Weeks DE, Matthews AG, Mukhopadhyay I. [2008] Dopaminergic mutations: within-family association and linkage in multiplex alcohol dependence families. *Am J Med Genet B Neuropsychiatr Genet* 147B (4):517-26.

[29] Berggren U, Fahlke C, Berglund KJ, Wadell K, Zetterberg H, Blennow K, Thelle D, Balldin J. [2010] Dopamine D2 receptor genotype is associated with increased mortality at a 10-year follow-up of alcohol-dependent individuals. *Alcohol Alcohol* 45:1-5.

[30] Kraschewski A, Reese J, Anghelescu I, Winterer G, Schmidt LG, Gallinat J, Finckh U, Rommelspacher H, Wernicke C. [2009] Association of the dopamine D2 receptor gene with alcohol dependence: haplotypes and subgroups of alcoholics as key factors for understanding receptor function. *Pharmacogenet Genomics* 19: 513-527.

[31] Schoots O, Van Tol HH. [2003] The human dopamine D4 receptor repeat sequences modulate expression. *Pharmacogenomics J* 3: 343-348.

[32] Oak JN, Oldenhof J, Van Tol HH. [200] The dopamine D(4) receptor: one decade of research. *Eur J Pharmacol* 405: 303-327.

[33] McGeary JE, Esposito-Smythers C, Spirito A, Monti PM. [2007] Associations of the dopamine D4 receptor gene VNTR polymorphism with drug use in adolescent psychiatric inpatients. *Pharmacol Biochem Behav* 86: 401-406.

[34] Lusher JM, Chandler C, Ball D. [2001] Dopamine D4 receptor gene (DRD4) is associated with Novelty Seeking (NS) and substance abuse: the saga continues. *Mol Psychiatry* 6: 497-499.

[35] Biederman J, Petty CR, Ten Haagen KS, Small J, Doyle AE, Spencer T, Mick E, Monuteaux MC, Smoller JW, Faraone SV. [2009] Effect of candidate gene polymorphisms on the course of attention deficit hyperactivity disorder. *Psychiatry Res* 170: 199-203.

[36] Grzywacz A, Kucharska-Mazur J, Samochowiec J. [2009] Association studies of dopamine D4 receptor gene exon 3 in patients with alcohol dependence *Psychiatr Pol* 42: 453-461.

[37] Kotler M, Cohen H, Segman R, Gritsenko I, Nemanov L, Lerer B, Kramer I, Zer-Zion M, Kletz I, Ebstein RP. [1997] Excess dopamine D4 receptor (D4DR) exon III seven repeat allele in opioid-dependent subjects. *Mol Psychiatry* 2: 251-254.

[38] Vandenbergh DJ. [1998] Molecular cloning of neurotransmitter transporter genes: beyond coding region of cDNA. *Methods Enzymol* 296: 498-514.

[39] Michelhaugh SK, Fiskerstrand C, Lovejoy E, Bannon MJ, Quinn JP. [2001] The dopamine transporter gene (SLC6A3) variable number of tandem repeats domain enhances transcription in dopamine neurons. *J Neurochem* 79:1033-8.

[40] Guindalini C, Howard M, Haddley K, Laranjeira R, Collier D, Ammar N, Craig I, O'Gara C, Bubb VJ, Greenwood T, Kelsoe J, Asherson P, Murray RM, Castelo A, Quinn JP, Vallada H, Breen G. [2006] A dopamine transporter gene functional variant associated with cocaine abuse in a Brazilian sample. *Proc Natl Acad Sci U S A* 103: 4552-7.

[41] Vandenbergh DJ, Bennett CJ, Grant MD, Strasser AA, O'Connor R, Stauffer RL, Vogler GP, Kozlowski LT.[2002] Smoking status and the human dopamine transporter variable number of tandem repeats (VNTR) polymorphism: failure to replicate and finding that never-smokers may be different. *Nicotine Tob Res* 4: 333-340.

[42] Cook EH Jr, Stein MA, Krasowski MD, Cox NJ, Olkon DM, Kieffer JE, Leventhal BL. [1995] Association of attention-deficit disorder and the dopamine transporter gene. *Am J Hum Genet*. 56: 993-998.

[43] Lee SS, Lahey BB, Waldman I, Van Hulle CA, Rathouz P, Pelham WE, Loney J, Cook EH. [2007] Association of dopamine transporter genotype with disruptive behavior

ISSN: 0976-3104

disorders in an eight-year longitudinal study of children and adolescents. *Am J Med Genet B Neuropsychiatr Genet* 144B: 310-317.

[44] Serý O, Didden W, Mikes V, Pitelová R, Znojil V, Zvolský P.[2006] The association between high-activity COMT allele and alcoholism. *Neuro Endocrinol Lett* 27: 231-235.

[45] Cao L, Li T, Liu X. [2003] Association study of heroin dependence and catechol-O-methyltransferase gene. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 20:127-130.

[46] Vandenbergh DJ, Rodriguez LA, Miller IT, Uhl GR, Lachman HM.[1997] High-activity catechol-O-methyltransferase allele is more prevalent in polysubstance abusers. *Am J Med Genet* 74: 439-442.

[47] Wang T, Franke P, Neidt H, Cichon S, Knapp M, Lichtermann D, Maier W, Propping P, Nöthen MM.[2001] Association study of the low-activity allele of catechol-O-methyltransferase and alcoholism using a family-based approach. *Mol Psychiatry* 6:109-111.

[48] Samochowiec J, Kucharska-Mazur J, Grzywacz A, Jabłoński M, Rommelspacher H, Samochowiec A, Sznabowicz M, Horodnicki J, Sagan L, Pełka-Wysiecka J.[2006] Family-based and case-control study of DRD2, DAT, 5HTT, COMT genes polymorphisms in alcohol dependence. *Neurosci Lett* 410: 1-5.

[49] Shih JC. [1991] Molecular basis of human MAO A and B. *Neuropsychopharmacology* 4: 1-7.

[50] Zhu QS, Shih JC. [1997] An extensive repeat structure down-regulates human monoamine oxidase A promoter activity independent of an initiator-like sequence. *J Neurochem* 69:1368-73.

[51] Brummett BH, Krystal AD, Siegler IC, Kuhn C, Surwit RS, Züchner S, Ashley-Koch A, Barefoot JC, Williams RB.[2007] Associations of a regulatory polymorphism of monoamine oxidase-A gene promoter (MAOA-uVNTR) with symptoms of depression and sleep quality. *Psychosom Med* 69: 396-401.

[52] Shih,JC, Chen, K and Ridd MJ [1999] Monoamine Oxidase: From Genes to Behavior. Annu Rev Neurosci. 22: 197–217.

[53] Lee SY, Hahn CY, Lee JF, Huang SY, Chen SL, Kuo PH, Lee IH, Yeh TL, Yang YK, Chen SH, Ko HC, Lu RB. [2010] MAOA Interacts With the ALDH2 Gene in Anxiety-Depression Alcohol Dependence. *Alcohol Clin Exp Res*. (in press)

[54] Brummett BH, Boyle SH, Siegler IC, Kuhn CM, Surwit RS, Garrett ME, Collins A, Ashley-Koch A, Williams RB.[2008] HPA axis function in male caregivers: effect of the monoamine oxidase-A gene promoter (MAOA-uVNTR). *Biol Psychol*79: 250-255.

[55] Lee SY, Hahn CY, Lee JF, Chen SL, Chen SH, Yeh TL, Kuo PH, Lee IH, Yang YK, Huang SY, Ko HC, Lu RB. [2009] MAOA-uVNTR polymorphism may modify the protective effect of ALDH2 gene against alcohol dependence in antisocial personality disorder. *Alcohol Clin Exp Res* 33: 985-990.

[56] Huang SY, Lin WW, Wan FJ, Chang AJ, Ko HC, Wang TJ, Wu PL, Lu RB. [2007] Monoamine oxidase-A polymorphisms might modify the association between the dopamine D2 receptor gene and alcohol dependence. *J Psychiatry Neurosci* 32:185-192.

[57] Vanyukov MM, Maher BS, Devlin B, Tarter RE, Kirillova GP, Yu LM, Ferrell RE. [2004] Haplotypes of the monoamine oxidase genes and the risk for substance use disorders. *Am J Med Genet B Neuropsychiatr Genet* 125B:120-125.

[58] Saiz PA, Garcia-Portilla MP, Florez G, Arango C, Corcoran P, Morales B, Bascaran MT, Alvarez C, San Narciso G, Carreño E, Alvarez V, Coto E, Bobes J.[2009] Differential role of serotonergic polymorphisms in alcohol and heroin dependence. *Prog Neuropsychopharmacol Biol Psychiatry* 33: 695-700.

[59] Seneviratne C, Huang W, Ait-Daoud N, Li MD, Johnson BA.[2009] Characterization of a functional polymorphism in the 3' UTR of SLC6A4 and its association with drinking intensity. *Alcohol Clin Exp Res* 33: 332-339.

[60] Conner BT, Hellemann GS, Ritchie TL, Noble EP. [2010] Genetic, personality, and environmental predictors of drug use in adolescents. *J Subst Abuse Trea*. 38:178-190.

[61] Bossert JM, Ghitza UE, Lu L, Epstein DH, Shaham Y.[2005] Neurobiology of relapse to heroin and cocaine seeking: an update and clinical implications. *Eur J Pharmacol* 526: 36-50.

[62] Li CY, Mao X, Wei L. [2008] Genes and (common) pathways underlying drug addiction. *PLoS Comput Biol* 2008; 4(1):e2. (http://karg.cbi.pku.edu.cn).

[63] Yang BZ, Kranzler HR, Zhao H, Gruen JR, Luo X, Gelernter J. [2008] Haplotypic variants in DRD2, ANKK1, TTC12, and NCAM1 are associated with comorbid alcohol and drug dependence. *Alcohol Clin Exp Res* 32: 2117-27.

[64] Miller D, Bowirrat A, Manka M et al. [2010] Acute Intravenous Synaptamine Complex [KB220] ™ Variant "Normalizes" Neurological Dysregulation in Patients during Protracted Abstinence from Alcohol and Opiates as Observed Using Quantitative Electroencephalographic (qEEG) and Neurotransmitter Genetic Analysis: Two Case Reports. *Postgraduate Medicine* (in press).

[65] Blum K ,Chen TJH, Morse S et al. [2010] Overcoming qEEG Abnormalities and Reward Gene Deficits During Protracted Abstinence In Male Psychostimulant and Polydrug Abusers Utilizing Putative Dopamine D2 Agonist Therapy. *Postgraduate Medicine* (in press).

[66] Vandenbergh, D. J., Perisco, A. M., Hawkins, A. L., Griffin, C. A., Li X., Jabs, E. W., and Uhl, G. R. [1992] Human dopamine transporter gene (DAT1) maps to chromosome 5p15.3 and displays a VNTR. *Genomics* 14:1104-1106.

[67] Anchordoquy HC, McGeary C, Liu L, Krauter KS and Smolen A. [2003]Genotyping of three candidate genes following whole genome preamplification of DNA collected from buccal cells. *Behav Genet* 33: 73-78.

[68] Van Tol, H.H., Wu, C.M., Guan, H.C., Ohara, K., Bunzow, J.R., Civelli, O., Kennedy, J., Seeman, P., Niznik H.B. and Jovanovic, V. [1992] Multiple dopamine D4 receptor variants in the human population. *Nature* 358: 149-152.

[69] Lerman, C., Caporaso, N., Main, D., Audrain, J., Boyd, N.R., Bowman, E.D. and Shields, P.G. [1998]. Depression and self-medication with nicotine: the modifying influence of the dopamine D4 receptor gene. *Health Psychol* 17: 56-62.

[70] Sabol S.Z., Hu, S. and Hamer D.[1998] A functional polymorphism in the monoamine oxidase A gene promoter. *Hum Genet* 103: 273-279.

[71] Caspi A., McClay J, Moffitt TE, Mill J, Martin J, Craig IW, Taylor A, Poulton R. [2002] Role of genotype in the cycle of violence in maltreated children. *Science* 297: 851-854.

[72] Haberstick, B.C., Lessem, J.M., Hopfer, C.J., Smolen, A., Ehringer, M.A., Timberlake, D., and Hewitt, J.K.[2005]

ISSN: 0976-3104

MAOA genotype and antisocial behaviors in the presence of childhood and adolescent maltreatment. *Am J Hum Genet* (Neuropsych. Genet). 135B: 59-64.

[73] Heils, A., A. Teufel, S. Petri, G. Stober, P. Riederer, D. Bengel, and K. P. Lesch. [1992] Allelic Variation of the Human Serotonin Transporter Gene Expression. *Journal of Neurochemistry* 66: 2621-2624.

[74] Lesch, K. P., Bengel, D., Heils, A., Sabol, S. Z., Greenberg, B. D., Petri, S., Benjamin, J., Muller, C. R., Hamer, D. H., and Murphy, D. L. [1996] Association of anxiety-related traits with a polymorphism in the serotonin transporter gene regulatory region. *Science* 274: 1527-1531.

[75] Hu, X., Oroszi, G., Chun, J., Smith, T.L., Goldman,D., and Schuckit, M.A.[2005] An expanded evaluation of the relationship of four alleles to the level of response to alcohol and the alcoholism risk Alcoholism: *Clinical and Experimental Research* 29: 8-16.

[76] Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS.[1991] 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res* 19: 4008.

[77] Pohjalainen T, Rinne JO, Nagren K, Lehikoinen P, Anttila K, Syvälahti EK, Hietala J. [1998] The A1 allele of the human D2 dopamine receptor gene predicts low D2 receptor availability in healthy volunteers. *Mol Psychiatry* 3: 256-260.

[78] Haberstick, B.C. and Smolen, A.[2004] Genotyping of three single nucleotide polymorphisms following whole genome preamplification of DNA collected from buccal cells. *Behav Genet* 34: 541-547.

[79] Mannisto PT, Kaakkola S. [1999] Catechol-O-methyltransferase (COMT): Biochemistry, Molecular Biology, Pharmacology, and Clinical Efficacy of the New Selective COMT inhibitors. *Pharm Rev* 51: 593-628.

[80] Akil M, Kolachana BS, Rothmond DA, Hyde TM, Weinberger DR, Kleinman JE. [2003]Catechol-O-methyltransferase genotype and dopamine regulation in the human brain. *J Neurosci* 23: 2008-2013.

[81] Bowirrat A, Oscar-Berman M. [2005] Relationship between dopaminergic neurotransmission, alcoholism, and Reward Deficiency syndrome. *Am J Med Genet B Neuropsychiatr Genet* 132B(1): 29-37.

[82] Giordano J, Blum K. [2010] Probing the Mysteries of Recovery through Nutrigenomic and Holistic Medicine: "Science Meets Recovery" through Seminal Translational Research. *Counselor Magazine* (in press)

[83] Blum K, Chen AL, Chen TJ, Braverman ER, Reinking J, Blum SH, Cassel K, Downs BW, Waite RL, Williams L, Prihoda TJ, Kerner MM, Palomo T, Comings DE, Tung H, Rhoades P, Oscar-Berman M. [2008]Activation instead of blocking mesolimbic dopaminergic reward circuitry is a preferred modality in the long term treatment of reward deficiency syndrome (RDS): a commentary. *Theor Biol Med Model* 12: 5-24.

[84] Kirsch P, Reuter M, Mier D, Lonsdorf T, Stark R, Gallhofer B, Vaitl D, Hennig J. [2006] Imaging gene-substance interactions: the effect of the DRD2 TaqIA polymorphism and the dopamine agonist bromocriptine on the brain activation during the anticipation of reward. *Neurosci Lett* 405:196-201.

[85] Stice E, Yokum S, Bohon C, Marti N, Smolen A. [2010] Reward circuitry responsivity to food predicts future increases in body mass: moderating effects of DRD2 and DRD4. *Neuroimage* 50: 1618-25.

[86] Volkow ND, Wang GJ, Begleiter H, Porjesz B, Fowler JS, Telang F, Wong C, Ma Y, Logan J, Goldstein R, Alexoff D, Thanos PK.[2006] High levels of dopamine D2 receptors in unaffected members of alcoholic families: possible protective factors. *Arch Gen Psychiatry* 63: 999-1008.

[87] de Courten-Myers GM.[1999] The human cerebral cortex: gender differences in structure and function. *J Neuropathol Exp Neurol* 58: 217-226.

[88] Kirisci T, Tarter RE, Vanyuku M, et al. [2006] Application of intern response theory to quantify substance use disorder severity. *Addictive behaviors* 31: 1035-1049.

[89] Blum K, Chen TJH, Bowirrat A et al. [2009] Genes & Happiness. *Gene Therapy and Mol. Biol* 13: 82-120.

[90] Blum K, Kozlowski GP. [1990] Ethanol and neurotransmitter interactions: a cascade model of reward. In: Ollat H, Parvez S, Parvez H. eds. *Alchol and Behavior, Uterecht,The Netherlands:VSP Press*: 131-149.

The IIOAB Journal

integrative omics

DATA ANNOTATION AND RELATIONS MODELING FOR INTEGRATED OMICS IN CLINICAL RESEARCH

Arno Lukas[1], and Bernd Mayer [1,2,*]

[1] *emergentec biodevelopment GmbH, Rathausstrasse 5/3, 1010 Vienna, Austria*

[2] *Institute for Theoretical Chemistry, University of Vienna, Waehringerstrasse 17, 1090 Vienna, Austria*

Received on: 03-April-2010; **Revised on:** 19-May-2010; **Accepted on:** 21-June-2010; **Published on:** 22-June-2010.

[*]**Corresponding author:** Email: bernd.mayer@emergentec.com Tel: +43-14034966; Fax: +43-14034966

ABSTRACT

Omics has massively permeated translational clinical research with numerous diseases being covered by Omics studies from the genome to the metabolome level. Integrating these disease specific Omics tracks appears a logical next step for building the fundament of Systems Biology and Systems Medicine. Here, coherence of individual Omics tracks regarding clinical hypothesis, samples and clinical descriptors, and finally data handling and integration become pivotal. We present a data integration, annotation and relations modeling concept for heterogeneous Omics data and workflows. With molecular features at the center of all Omics we link the result profiles from different Omics tracks characterizing a specific disease phenotype to a common human molecular reference network for allowing a seamless integration and subsequent support in interpretation of Omics screening results.

Our concept rests on data structures for representing objects specified by metadata and content. For handling diverse Omics tracks a flexible structure for content is proposed allowing data representation at different levels of granularity as demanded by the type of Omics and specific type of data. Content on the molecular level includes deep annotation of molecular features on gene and protein level. Based on this annotation pair-wise relations between molecular objects are, traversing the molecular annotation into a network of relations (molecular feature graph). Such a relation network is also built on the Omics data level, combining explicit relations derived from study setup and implicit relations generated by mining metadata and content (Omics data graph).

Finally both graphs are merged utilizing the molecular feature level as common denominator, enabling a persistent integration and subsequently interpretation of Omics profiling results in the realm of a given clinical hypothesis. We present a case study on integrating transcriptomics and proteomics data on chronic kidney disease for demonstrating the feasibility of this concept.

.

Keywords: *integration; networks; standards; omics; graphs*

[I] INTRODUCTION

With sequencing of the human genome a major cataloguing milestone was reached in 2001 [1], followed by rapid development of Omics tracks spanning from the genome to the metabolome level. A summary statistics on the various Omes is provided at the Gerstein lab (http://bioinfo.mbb.yale.edu/what-is-it/omes/omes.html), clearly indicating the maturity of genomics efforts when compared to the other Omes. Omics has in the meantime entered clinical sciences aimed at elucidating the pathophysiology of diseases, thereby providing the basis for identifying biomarkers serving for novel diagnostics and therapy [2,3]. Specific profiling has already been forwarded to clinical application, e.g. for assessing breast cancer utilizing a profile of about 70 features [4]. Numerous prevalent diseases have been studied on the various Omics levels, and first efforts were introduced for consolidating this body of knowledge in open access data repositories. Usually these repositories are Omics-specific as e.g. ArrayExpress for transcriptomics [5] or PRIDE for proteomics data [6], or Omics profiles are consolidated on the level of genes (gene-centric) as in Genecards (http://www.genecards.org) [7]. For some etiologies also disease-specific databases have been established, with Oncomine as an example for consolidating cancer transcriptomics data [8]. Platforms integrating various Omics levels, however, are less common, although being perfectly in

line with approaches in Systems Biology [9], in the meantime already expanding at least conceptually towards Systems Medicine [10]. Aim of these concepts is broad integration of Omics tracks being embedded in clinical data space and sample descriptors, with the ultimate goal of providing a quantitative representation of disease (outcome)- specific molecular processes.

Distinct specifications have to be met in Omics in particular including: i) a quantitative assessment of molecular objects, and ii) approaching the totality of objects at some layer of cellular organization. Advancements in miniaturization, improved readout technologies, and parallelization of established technologies have significantly contributed to the accuracy of quantitative measurement procedures. However, major shortcomings remain with cataloguing efforts for determining the totality of some sort. Here, genomics may come closest to completeness, presently experiencing a further boost resting on next generation sequencing technologies at least in principal allowing an unbiased decoding of entire genomes [11]. However, for all other Omes limitations have to be recognized, and even the notion of a "gene" came under some scrutiny, [12] particularly when evaluating results of the ENCODE consortium [13]. Gene expression array data in most cases still focus on protein coding genes, may include some resolution on the level of splice variants, but only in rare cases expand to assessing miRNAs or more generally ncRNAs. The totality of the proteome (and to some extent also of the metabolome) is under question on a theoretical level, but is rapidly evolving due to parallelized high resolution separation, identification, as well as quantification.

For integrative Omics, and here in particular in the medical context, numerous additional factors have to be taken into consideration, centrally including sample specifications [14]. A detailed clinical hypothesis comes in the first place, and from there delineation of strict sample inclusion and exclusion criteria result. Case-control studies are the typical setup in screening, where ideally cases and controls are matched for all parameters with known or suspected impact but the clinical question of interest (outcome). Here either a dedicated prospective sample and data collection has to be established, or a retrospective collection is available. Best sources in the latter case include interventional studies performed under strict quality control. In line with sample specification is assessment of sample size for assuring a well powered study from the statistical perspective for each individual Omics track considered [15]. Omics procedures are applicable for various sample types, most frequently utilizing tissue, blood and urine. Here standardized sample handling and preparation comes into play, where standard operating procedures (SOPs) for storage and preparation have been derived for a number of Omics tracks [16].

In the light of the aforesaid the following issues may be considered as central for integrating heterogeneous Omics profiling results:

1. Thorough definition of the clinical hypothesis
2. Detailed specification of cases and controls for each Omics track
3. Sample size calculations for each specific Omics track
4. SOPs for sample and clinical data handling
5. SOPs for Omics procedures and data generation
6. Standardized reporting covering each Omics workflow

Regarding reporting conventions numerous initiatives have been started, including experiment description as well as execution standards [17], and both are to some extend already followed in results reporting, with MIAME being a well known implementation for transcriptomics [18].

If different Omics tracks follow defined *standards* in reporting and are *in line with a given clinical hypothesis* Omics integration on the level of result profiles becomes feasible. For setting up a cross-Omics results integration two approaches may be followed for data preparation: Public domain driven by consolidating available information on a given clinical hypothesis (e.g. by extracting available profiles on a specific disease from ArrayExpress of PRIDE), or implementation of a dedicated cross-Omics project explicitly focusing on the specific clinical question. The latter approach may even expand towards using samples from the very same patients for conducting the individual Omics tracks, certainly adding to data coherence. Prototypical settings of such initiatives include the research consortia predict-IV focusing on toxicological aspects (http://www.predict-iv.toxi.uni-wuerzburg.de), or SysKid (http://www.syskid.eu) analyzing chronic kidney disease by a Systems Biology approach.

Fulfillment of the technological and procedural requirements discussed so far enable consolidation of heterogeneous Omics feature profiles in a Systems Biology (Medicine) context. The next step in implementing such an approach is providing data management and integration which serves as basis for subsequent analysis, ultimately yielding molecular processes, biomarkers and target candidates linked to the specific disease and outcome. At this step the incomplete molecular cataloguing aspect comes in, adding *annotation* as a major aspect to Omics data management and integration.
We in the following propose a data consolidation and annotation framework specifically aimed at covering integration of diverse Omics result profiles directly linked to a human molecular reference network. We in particular present concepts for explicit as well as implicit relations inference aimed at supporting data interpretation in the realm of a given clinical hypothesis.

[II] MATERIALS AND METHODS

2.1. Object abstraction

The generic component of our concept is an *object*, resembling a data structure holding a unique identifier. Practical notion of an object is kept broad, involving molecular objects and Omics data objects. Omics data objects, in the following referred to as "records", involve any type of machine readable data relevant for or generated in the course of experimental procedures. Typical records include raw data matrices,

analysis results (being the core of our integration concept), validation results, or sample specifications. Molecular objects on the other hand are defined as known and well annotated genes or proteins (but conceptually may be expanded for also including RNA, metabolites, etc.). For each object metadata are provided allowing further characterization of the object category. Next, the effective *content* of an object is given. Molecular content involves annotation data e.g. specifying a gene's functional terms or protein interaction data. Omics record content is in a first place characterized by the level of granularity, where content of an individual record may involve large profiling matrices covering an entire Omics screening experiment, may resemble results profiles from case-control studies, or may provide individual molecular features and their specific expression value found in a particular experiment. A third major element is *relations* which put objects (and their content) into context. Relations again follow the data structure concept, where next to a unique identifier metadata are provided. Relation specific metadata mainly include a specification of the type and further edge content as directionality, source (explicitly built or implicitly computed), or evidence level.

2.2. Technical implementation

The Java Enterprise Platform (http://java.sun.com/javaee), utilizing a post-relational approach as data foundation, provides an efficient platform for implementing object oriented concepts as discussed here. This platform supports dynamic data models technically enabled via the Content Repository for Java (JCR), complemented by Glassfish as application server. On the server side the Enterprise Java Bean component architecture seamlessly supports an architectural design for separating application logic and presentation logic. Apache Jackrabbit as a reference implementation of JCR provides further functionality including versioning and full text search. Java Server Faces may be used for implementing the client side.

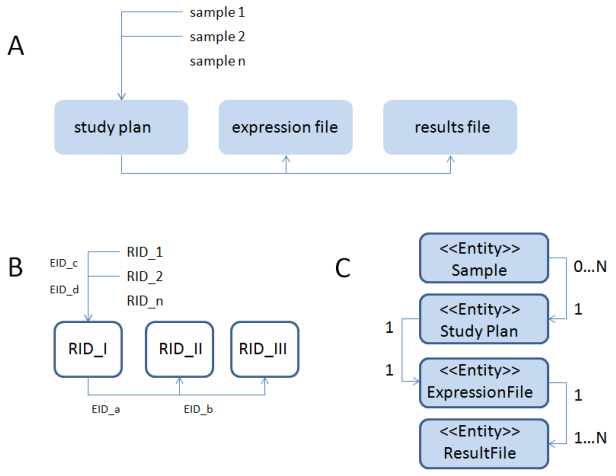
2.3. Public domain sources

Software platforms, modules, as well as molecular content necessary for realizing the technical backbone of the concept presented in this work are available in the public domain. The JCR reference Apache Jackrabbit is found at http://jackrabbit.apache.org, Java Server Faces is found at http://java.sun.com/javaee/javaserverfaces. A manifold of modules for supporting data processing workflows is provided by Taverna (http://www.taverna.org.uk), with the Taverna engine also embedded in Java. Biomart, available at http://www.biomart.org, can be customized for supporting the data management side, and additionally a website can be configured for providing user interfaces. Biomart further allows interfacing via web services for handling large data sets. As objects are represented in their context visualization of resulting networks is essential for supporting interpretation. Gehlenborg et al. [9] recently provided a review on Omics visualization tools, with Cytoscape (http://www.cytoscape.org) as a prominent example. Cytoscape allows an extended definition and display of node (record) types, necessary for visualizing heterogeneous content spanning from clinical sample nodes to molecular feature nodes. Different types of molecular interaction networks are available for download, including procedural interactions from KEGG (http://www.genome.jp/kegg) and PANTHER (http://www.pantherdb.org), physical interactions (both experimentally determined as well as predicted) from the meta-database OPHID (http://ophid.utoronto.ca/ophidv2.201), or interaction networks consolidated from multiple sources as STRING (http://string.embl.de). For assuring coherence on the name space level for molecular reference networks as well as molecular features coming from the various Omics levels a reference namespace has to be selected and regularly updated. Source providing broad coverage of features are found with UNIPROT (http://www.uniprot.org) or NCBI (http://www.ncbi.nlm.nih.gov/refseq).

[III] RESULTS

3.1. Omics record consolidation

The generic object for Omics data consolidation is a record representing data at any given level of detail, e.g. characterizing an entire transcriptomics profile or only a single feature and its associated expression value. For each record metadata may be provided for further characterization of the record content. Furthermore object relations can be built for introducing dependencies between records. **[Figure-1]** provides an example scheme of the record (node) and relation (edge) concept.



Omics workflows: (A) Schematic setup of an Omics track involving study plan, expression raw data and analysis results data. (B) Formal representation of the workflow as node and edge concept with each object encoded as a data structure holding a unique identifier and a parameter list (C) Representation of the concept in UML (Unified Modeling Language, http://www.uml.org).

Omics procedures follow a generic process as exemplified in Figure 1A. First a study plan is specified defining the case-control setup reflecting the clinical hypothesis, methodology used, etc. Based on the definition of cases and controls samples are linked which effectively undergo screening as specified in the study plan. Equivalent to study plans samples are represented as records (holding sample source, type, amount available, etc. as metadata and content). Frequently samples are organized in dedicated databases and may only be linked into an Omics record management via unique sample identifiers. When retrieving Omics profiles from the public domain the level of detail regarding sample-specific clinical data is frequently sparse and typically limited to clinical categories/stages for the disease.

Executing experimental profiling results in an expression record (e.g. raw data matrix of case and control samples), which after statistical analysis leads to a results record only listing significantly differentially regulated features when comparing case and control group. Typically such a results profile is based on a per-feature statistical test including correction for multiple testing. Although substantial differences in experimental procedures are evident this basic workflow is followed by most Omics tracks assessing continuous concentration values (a fact also becoming evident when comparing MIAME and MIAPE for transcriptomics and proteomics, respectively).

On an abstract level (Figure 1B) a graph representation becomes feasible, holding nodes characterized by record identifiers (RID) and edges specified by edge identifiers (EID, in the example case being directed). Each node and each edge is accompanied by a data structure holding a unique identifier. In the case of nodes metadata and the content as such are stored in the data structure, for edges the node identifiers specifying the connectivity via node IDs as well as metadata (directionality, type of edge, etc.) are provided. For nodes individual content may be represented at arbitrary levels of granularity (spanning from whole profile matrices to single features) depending on subsequent resolution needs in analysis. However, resolution on the level of individual features is mandatory in virtually all analysis procedures. For practicability issues encapsulation of entire profiles, arrays of profiles, or analysis result vectors appears preferable. This approach significantly reduces complexity on the record level and eases upload and management, but still provides access to individual features when using record templates (where e.g. feature and associated expression value reside in defined content locations).

Omics integration naturally demands a combination of profiling efforts, exemplarily shown in **[Figure-2]**. The situation given in Figure 2A is defined by individual study plans I-III, respective screening profiles (e.g. raw data) and results (list of significantly different features on the transcriptome, proteome and metabolome level). Multiple result files may be generated (see also the UML in Figure 1C) e.g. by varying statistical procedures used for analyzing a given case-control group, or by varying the assignment of samples as case and control.
In an ideal setting the studies are fed from a single sample / clinical descriptor repository (feasible for explicitly designed cross-Omics), or have to be extracted to the extent possible if fed from public domain Omics profiling (e.g. gathering available Omics studies regarding a specific clinical hypothesis). Naturally, a dedicated study will provide a more complete and coherent set of records, as these reflect explicitly defined inclusion criteria focusing on a specific clinical hypothesis.

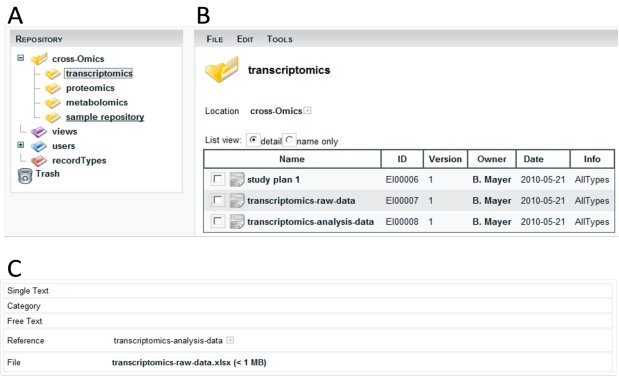


Fig: 2. Node and edge concept for cross-Omics: (A) Schematic setup of three individual Omics tracks fed by two sample sources. (B) Formal representation of the Omics tracks given in (A) further including two implicit relations (EID_1, EID_2).

Certainly, record types are not limited to the examples given in Figure 1 and 2. Further record types of value include scientific references, standard operating procedures, or experimental validation data (including both, profile validation as well as complementary data e.g. coming from *in-vitro* and *in-vivo* models of the clinical setting), among others.

[Figure-3] provides an implementation example for organizing the corresponding record management. This reference implementation organizes records along specific Omics tracks (Figure 3A), and essential records specifying the study specifications and results (Figure 3B). For each record metadata as well as explicit links between records (in the example linking transcriptomics raw data and analysis data, Figure 3C) can be specified. All relations specified in Figure 1A and 2A are *explicit*, as such defined by the user depositing the records, and reflect the logical structure of Omics procedures. Of central relevance here is that the Omics tracks are driven as independent processes, in a first place only (explicitly) linked if using joint samples (and more generic by focusing on one specific clinical hypothesis).

However, further *implicit* relations are present in the collection of records (Figure 2B). One set of relations may be derived from joint metadata (EID_1) used for characterizing records (e.g. using the same tissue type), and a second set of relations may be derived from the record content as such (EID_2): Software frameworks as Jackrabbit provide full indexing of records for text search, and by this mechanism records can be linked e.g. based on “overall similarity”, or specifically by e.g. invoking on joint molecular identifiers in feature lists. Relevant examples include shared gene or protein identifiers for extracting relations e.g. between transcriptomics and proteomics profiles.

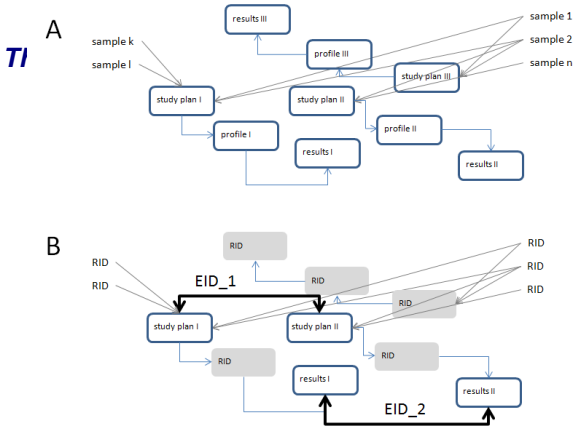


Fig: 3. Example layout for cross-Omics record management: (A) Repository structure involving three Omics tracks and a clinical sample repository. (B) Records assigned to a specific Omics track (see Figure 1A). (C) Example metadata categories for a transcriptomics raw data record including an explicit relation between transcriptomics raw data and analysis data.

Annotation of relations certainly goes far beyond the examples provided here, as numerous project specific term lists may be used. Relevant examples include relations mapping based on disease-feature or feature-drug associations: A results file record may be mined for occurrence of gene or protein identifiers with known links to diseases (e.g. utilizing OMIM data, http://www.ncbi.nlm.nih.gov/omim), and if identified the diseases can be added as metadata to the record. In a next step relations between records can be built based on co-occurrence of disease associations. A comparable procedure may be relevant for known drug-target associations as e.g. provided by STITCH [19]. Yet another relevant procedure is to link scientific publications to records via publication-feature information e.g. mined from MEDLINE [20].

3.2. Omics feature consolidation

Equivalent to the graph concept for representing Omics workflows also molecular features can be consolidated. In a standard Omics setup a feature denotes a relevant object (gene, transcript, protein, etc.) separating cases and controls utilizing a statistical measure. The typical representation of features including their relations is graphs, with protein-protein interaction networks (PPIs) as well known example [21]. PPIs are usually specific regarding the type of relation, e.g. IntAct networks encode physical (undirected) interactions [22], whereas KEGG represents procedural information also including edge directionality [23]. We derived the human proteome interaction network omicsNET which combines significant annotation with relations modeling. RefSeq (http://www.ncbi.nlm.nih.gov/refseq) is used as reference source for human genes and proteins providing about 25,000 objects (considering a canonical sequence set of genes and proteins). For each gene/protein deep annotation was performed utilizing public domain sources, including tissue specific reference gene expression, various sources for functional annotation as Gene Ontologies, manifold protein interaction data sources and further protein characterization as subcellular location, among others. Additionally transcriptional control on the level of transcription factors and miRNAs was added.

Technically, data structures were used, each holding a unique identifier linking to a gene/protein, and storing the annotation data as content. On the basis of the gene/protein-specific content a pair-wise relation score was computed which may be interpreted as dependency resting on the individual annotation given in the content. For further details on omicsNET we refer to [24]. A schematic layout of the construction principle is given in **[Figure-4]**.

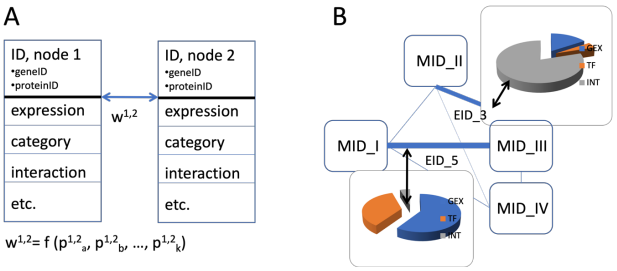


Fig: 4. Concept of a molecular feature annotation graph: (A) Data structures specified by unique identifiers and extended annotation serve as basis for computing dependency weights. (B) Network representation holding molecular nodes and weighted edges, where weights are delineated from given annotation exemplarily shown for two edges (EID_3, EID_5) (individual contributions coming from: GEX: tissue specific gene expression; TF: joint transcription factor binding sites; INT: protein-protein interactions).

Next to consolidated annotation the feature representation given in Figure 4 provides the opportunity for automated relations modeling. All content associated with molecular nodes is parameterized as input for an empirical metafunction *f* which allows computing pair-wise dependencies between molecular features. The metafunction integrates similarity measures as correlation coefficients for tissue specific gene expression profiles, as well as dependency measures as known protein interaction (e.g. coming from Intact or KEGG) for a given pair. The resulting parameter $w^{x,y}$ approximates an aggregate dependency between molecular features, and as this is done for all features a complete matrix and graph results. This graph can now be used for mapping analysis results coming from the various Omics tracks.

3.3. Integrating records and features

Obviously the representation of multiple Omics workflows, but also the system analyzed by Omics as such, namely an extended (although far from being complete) assessment of molecular entities may be represented as nodes (content) and edges (relations).

On this basis an integration of both structures is an obvious next step, as schematically shown in **[Figure-5]**.

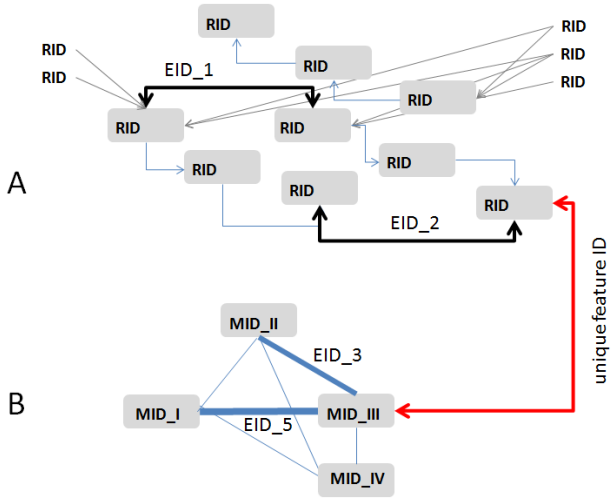


Fig: 5. Integrating Omics results data on a molecular feature graph: (A) Record data structures model (see also Figure 2B) and (B) feature data structure (see also Figure 4B) interlinked on a joint name space level (edge given in red).

Omics operates on molecular name spaces, with gene and protein IDs as the most prevalent reference spaces. Decomposing all data into records with unique identifiers naturally supports building relations between the Omics record structure and the molecular feature structure. From this concept a persistent relations mapping for Omics results integration emerges, embedding sample space, experimental procedure logics, and molecular feature landscape. Features identified as relevant on the record level (stored in Omics result records) have a direct representation on the feature graph and vice versa.

As for all relational models querying is naturally supported by the presented concept. However, yet another more powerful type of querying becomes feasible, namely subgraph extraction. An example subgraph is schematically depicted in **[Figure-6]**.

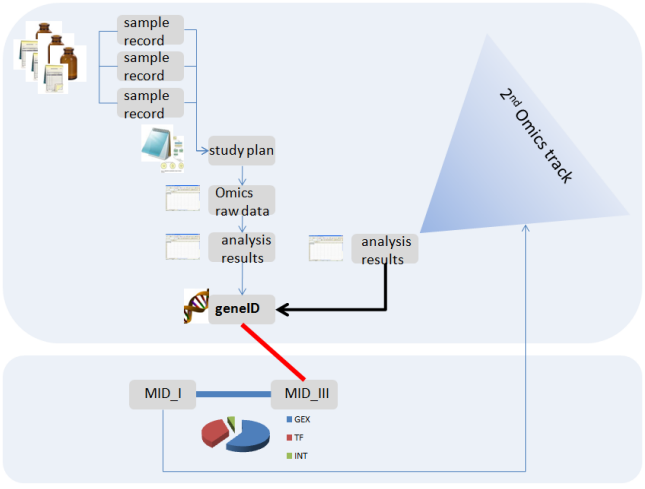


Fig: 6. Navigating in Omics record and molecular feature space: For a particular Omics track explicit relations are provided from sample records to study plan, further to raw data and results data. A specific feature of interest (geneID) given in results data shows an implicit link to a results record from a second Omics track, and on the molecular level equivalency with MID_III, which further shows strong dependency to MID_I (and may link to a result profile coming from a different Omics track).

Merging the record and feature concepts traverses the traditional querying in relational databases into analysis of subgraphs. The example provided in Figure 6 uses a particular feature from an Omics analysis results file as start point. For this feature an implicit link to a second results file coming from a different Omics track is detected which allows tracking the path upstream this second track.
At the same time downstream analysis into the molecular feature space becomes amenable. Here relations rest on computed dependencies based on broad feature annotation. For the example case a strong link to a second molecular feature may be followed which itself eventually may have become evident at some other level (e.g. an associated scientific publication) in a second Omics track.

3.4. Example case

We in the following exemplify the presented concept for Omics profiling of chronic kidney disease (CKD), a disease characterized by progressive loss of kidney function. CKD has been extensively studied on various Omics levels with an impressive consolidation effort on the transcriptomics level provided by the nephromine database (http://www.nephromine.org). Next to diabetic nephropathy (DN) and hypertensive nephrosclerosis other (mainly histopathological) classifications characterize the types of CKD, including IgA nephritis, focal segmental glomerulosclerosis, membraneous glomerulonephritis, and minimal change disease. For these types of CKD specific profiles on transcript and proteome level are available in the public domain, all derived on disease-type specific case-control Omics profiling [25]. Utilizing the cross-Omics integration concept outlined above provides a graph shown in [Figure 7].

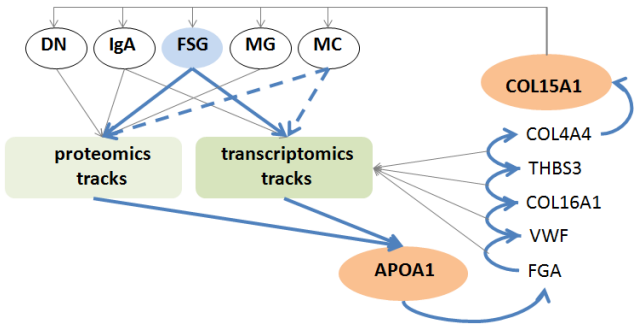


Fig: 7. Integrated Omics for chronic kidney disease: Sample classification is provided by the histopathological representation of the disease (DN: diabetic nephropathy; IgA: IgA nephritis; FSG: focal segmental glomerulosclerosis; MG: membraneous glomerulosclerosis; MC: minimal change disease), each type entering proteomics, transcriptomics, or both. Strong arrows indicate paths linking sample type, Omics track, and molecular feature space (given by gene symbols).

In this example Omics profiling is included for various types of CKD, with feature lists from proteomics characterizing all five representations when compared to matched (healthy) controls, and three conditions characterized by transcriptomics. Implicit linking of Omics result profiles shows APOA1 jointly identified by proteomics and transcriptomics when only considering the CKD type FSG. Analyzing APOA1 on the level of the protein interaction networks a path including the molecular identifiers FGA, VWF, COL16A1, THBS3 and COL4A4 and COL15A1 becomes evident (and all of these are identified as significantly differentially expressed by the transcriptomics track), where COL15A1 is additionally identified as significantly affected for all types of CKD on the basis of proteomics screening results. Consequently, also minimal change disease links into this network. Interesting to note here is that from a clinical perspective minimal change disease presents comparable to prolonged segmental glomerulosclerosis.

This Omics results annotation may be further extended by including genetic studies on CKD [26] identifying uromodulin (UMOD) as affected. UMOD itself is found as differentially regulated by the transcriptomics studies, and shows on the molecular graph level a shortest path to APOA1 (via CRP and APOA2), but also to COL15A1 (via CRP, FN1, and COL5A1).

[IV] DISCUSSION

Omics procedures have reached a level of maturity enabling implementation in standard laboratories, and broad scale application is seen in translational and clinical research. Standards have been derived for most Omics tracks including both experimental design as well as execution, and reproducibility of Omics screening shows satisfactory results. However, integration of results from different Omics tracks and domains, but even of results coming from Omics studies focusing on the very same level of molecular organization experience shortcomings. We consider two main issues as relevant. The first is maintaining strict coherence on the experimental side in particular regarding sample inclusion and processing criteria. Specifically when addressing complex situations as human diseases a strict definition of the clinical hypothesis, associated clinical parameters, and outcome have to be closely shared for individual Omics tracks aimed for integration. For illustrating this issue the clinical presentation of “chronic kidney disease” [27] may be used, which as term includes various causative conditions and on the level of outcome may involve various parameters as levels of albuminuria, creatinine, or glomerular filtration rate. Omics integration for “chronic kidney disease” will certainly provide a far less coherent picture on the molecular level than using studies addressing specific type and specific stage of the disease. Omics procedures following such strict inclusion are certainly less frequently found in the public domain emphasizing the importance of dedicated Omics approaches.

The second major issue is data handling concepts supporting Omics workflows on the entire level of annotation, spanning from the clinical data spectrum to the individual Omics profiles and relevant features resulting from the manifold of different analysis procedures. As mentioned above disease-specific Omics repositories slowly emerge, also including to some extent metadata information as sample specifications on the clinical level. However, most of presently found disease specific repositories in the public domain are too broad in scope, hamper metadata at an adequate level of detail, and mostly include only a specific Omics domain (with transcriptome profiles as the most abundant type).

We in this work present an Omics integration concept covering both, the data spectrum of Omics tracks as well as persistent mapping to molecular annotation. Data management concepts for Omics in a first place need a specification regarding granularity of data representation. Laboratory Information Management Systems (LIMS) have been designed for also covering Omics [28]. However, from the background of LIMS significant standardization of workflows is assumed which for individual Omics tracks appears manageable but for cross-Omics is difficult to maintain (and for repositories built from public domain is merely impossible to achieve). For handling this issue we propose a record concept, formally represented as data structure managing content at arbitrary levels of granularity, where templates serve for standardizing experimental design and execution. This data encapsulation provides easy adaption to expanding scope (e.g. if yet another Omics track becomes available and needs integration), but also allows a representation of the entire Omics workflow including study plans, sample repositories, procedure documentation, raw data files, as well as analysis results and verification data.
The proposed Omics annotation concept takes, next to data representation, care of another central aspect, namely relations modeling. Uniquely referenced objects allow explicit

definition of relation (as raw data file and associated analysis file(s)), and if implemented in a proper environment provides implicit relations modeling. The latter is of particular relevance on the level of cross-Omics data interpretation.

The combination of Omics procedure annotation and relations modeling traverses the concept into a knowledge representation framework, formally represented as graph with content (nodes) in their context (edges). Such a design naturally enables integration with molecular graphs with genes/proteins being the predominant levels for data interpretation (where e.g. metabolites are mapped to involved enzymes, or SNP data to affected genes including their regulatory regions). Various molecular graphs resting on deep annotation have been derived with omicsNET [24] or STRING [29] as prototypical reference. Merging Omics graphs and molecular graphs enables extended querying utilizing methodologies provided by graph theory [30]. The concept discussed above allows extracting subgraphs and paths linking molecular features to their neighborhood on the molecular, the Omics tracks, and the sample specifications (clinical) level.

[V] CONCLUSION

Omics integration clearly bears the potential of expanding our understanding of complex diseases, and substantial efforts for bridging Omics levels have already been reported [2,9,10,25,31]. However, for building descriptive models characterizing diseases at the interface of clinical specifications and molecular processes in the realm of higher order structures as proposed for formal and biochemical systems [32] more fundamental issues have to be tackled. We consider annotation and relations modeling embedded in flexible data and knowledge management frameworks as a fundament for concise cross-Omics data interpretation on the level of descriptive graphs. Only as cataloguing efforts on the molecular level expand, and the number of different Omics screens on specifically defined clinical etiologies increase, model building in the realm of Systems Biology and Systems Medicine will become amenable.

FINANCIAL DISCLOSURE

This work was financially supported by the European Union FP7 projects predict-IV (project number 202222) and SysKid (project number 241544).

REFERENCES

[1] International Human Genome Sequencing Consortium. [2001] Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

[2] Perco P, Wilflingseder J, Bernthaler A, et al. [2008] Biomarkers for cardiovascular disease and bone metabolism disorders in chronic kidney disease. A systems biology perspective. *J Cell Mol Med* 12:1177–87.

[3] Sikaroodi M, Galachiantz Y, Baranova A. [2010] Tumor markers: the potential of "omics" approach. *Curr Mol Med* 10:249–57.

[4] Slodkowska EA, Ross JS. [2009] MammaPrint 70-gene signature: another milestone in personalized medical care for breast cancer patients. *Expert Rev Mol Diagn* 9:417–22.

[5] Parkinson H, Kapushesky M, Kolesnikov N, et al. [2009] ArrayExpress update--from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res* 37:D868–72.

[6] Jones P, Martens L. [2010] Using the PRIDE proteomics identifications database for knowledge discovery and data analysis. *Methods Mol Biol* 604:297–307.

[7] Rebhan M, Chalifa-Caspi V, Prilusky J, et al. [1997] GeneCards: integrating information about genes, proteins and diseases. Trends Genet 13:163.

[8] Rhodes DR, Yu J, Shanker K, et al. [2004] ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* 6:1–6.

[9] Gehlenborg N, O'Donoghue SI, Baliga NS, et al. [2010] Visualization of omics data for systems biology. *Nat Methods* 7:56–68.

[10] Ahn AC, Tewari M, Poon C-S, et al. [2006] The Clinical Applications of a Systems Approach. *PLoS Med* 3:e209.

[11] Mamanova L, Coffey AJ, Scott CE, et al. [2010] Target-enrichment strategies for next-generation sequencing. *Nat Methods* 7:111–8.

[12] Stadler PF, Prohaska SJ, Forst CV, et al. [2009] Defining genes: a computational framework. *Theory Biosci* 128:165–70.

[13] Rosenbloom KR, Dreszer TR, Pheasant M, et al. [2010] ENCODE whole-genome data in the UCSC Genome Browser. *Nucleic Acids Res* 38:D620–5.

[14] Taylor CF [2007] Progress in standards for reporting omics data. *Curr Opin Drug Discov Devel* 10:254–63.

[15] Mehta TS, Zakharkin SO, Gadbury GL et al. [2006] Epistemological issues in omics and high-dimensional biology: give the people what they want. *Physiol Genomics* 28:24–32.

[16] Morrison N, Cochrane G, Faruque N, et al. [2006] Concept of Sample in OMICS Technology. *OMICS* 10:127–37.

[17] Ball CA, Brazma A. [2006] MGED standards: work in progress. OMICS 10:138–44.

[18] Quackenbush J. [2009] Data reporting standards: making the things we use better. *Genome Med* 25:111.

[19] Kuhn M, Szklarczyk D, Franceschini A, et al. [2010] STITCH 2: an interaction network database for small molecules and proteins. *Nucleic Acids Res* 38:D552-6.

[20] Matos S, Arrais JP, Maia-Rodrigues J, et al. [2010] Concept-based query expansion for retrieving gene related publications from MEDLINE. *BMC Bioinformatics* 11:212.

[21] Alberghina L, Höfer T, Vanoni M. [2009] Molecular networks and system-level properties. *J Biotechnol* 144:224–33.

[22] Aranda B, Achuthan P, Alam-Faruque Y, et al. [2010] The IntAct molecular interaction database in 2010. *Nucleic Acids Res* 38:D525–31.

[23] Kanehisa M, Goto S, Furumichi M, et al. [2010] KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 38:D355-60.

[24] Bernthaler A, Muehlberger I, Fechete R, et al. [2009] A dependency graph approach for the analysis of differential gene expression profiles. *Mol Biosyst* 5:1720-31.

[25] Perco P, Mühlberger I, Mayer G et al. [2010] Linking transcriptomic and proteomic data on the level of protein interaction networks. *Electrophoresis* 31:1780-1789.
[26] Köttgen A, Pattaro C, Böger CA, et al. [2010] New loci associated with kidney function and chronic kidney disease. *Nat Genet* 42:376-84.
[27] Levey AS, Astor BC, Stevens LA, et al. [2010] Chronic kidney disease, diabetes, and hypertension: what's in a name? *Kidney Int* 78:19-22.
[28] Wishart DS. [2007] Current progress in computational metabolomics. *Brief Bioinform* 8:279-93.
[29] von Mering C, Jensen LJ, Kuhn M, et al. [2006] STRING 7--recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* 35:D358-62.
[30] Platzer A, Perco P, Lukas A, et al. [2007] Characterization of protein-interaction networks in tumors. *BMC Bioinformatics* 8:224.
[31] Nibbe RK, Koyutürk M, Chance MR. [2010] An integrative -omics approach to identify functional sub-networks in human colorectal cancer. *PLoS Comput Biol* 6:e1000639.
[32] Rasmussen S, Baas NA, Mayer B, et al. [2001] Ansatz for dynamical hierarchies. *Artificial Life* 7: 329–53.

ABOUT AUTHORS:

Arno Lukas, Managing Partner of emergentec biodevelopment GmbH, Vienna, Austria, holds a PhD in Biochemistry. Arno's main research interests involve translational research towards discovery and validation of novel biomarkers in human disease.

Bernd Mayer, Managing Partner of emergentec biodevelopment GmbH , Vienna, Austria, and Associate at the Institute for Theoretical Chemistry, University of Vienna, holds a PhD in Molecular Biology. Bernd's research interests cover bioinformatics concepts and methods.