

Value of liberal arts education in the era of precision medicine

Abstract

The Precision Medicine Initiative (PMI) aims to treat and prevent disease by accounting for individual variability in genes, environment, and lifestyle for each person. The initiative announced by President Obama aims to build the PMI Cohort Program that will compile a national research cohort of one million or more U.S. participants. Through technical advances in research capability for high throughput screening of genes/gene products, data sharing technology and informed consent policies that empowers patients, the PMI will usher in a new era of medicine in which researchers, providers and patients work together to develop individualized care. Paramount to successful delivery of this initiative will require trust and participation of the public and scientists being able to share data and protect the rights of the public. It is our contention that liberal arts colleges can fully participate in PMI by: 1. Taking advantage of open source and cloud platforms to access and analyze data obtained from both basic and clinical sciences; and 2. Providing an intellectually diverse student body to survey opinions and knowledge of modern genomic technologies in order to identify areas of public distrust and issues with informed consent arising from the implementation of these technologies. Furthermore, the curriculum at liberal arts colleges will enable deep discussion and exploration of moral, ethical, legal and scientific aspects of large scale initiatives such as PMI.

Keywords: survey, genetic information, bioinformatic, open source, cloud computing, moral, ethics, undergraduate education, bioconductor, shiny

Volume 3 Issue 5 - 2016

Griffin Reed, Sanjive Qazi

Biology Department, Gustavus Adolphus College, USA

Correspondence: Sanjive Qazi, Gustavus Adolphus College, 800W College Avenue, Saint Peter, Minnesota, USA, Tel | 507 933 6319, Email sqazi@gustavus.edu

Received: May 25, 2016 | **Published:** May 27, 2016

Abbreviations: PMI, precision medicine initiative; FDA, food and drug administration; DNA, deoxyribonucleic acid; RNA, ribonucleic acid; WGS, whole genome sequencing; GT, gene therapy; SYK, spleen tyrosine kinase; GEO, gene expression omnibus; EMBL, european molecular biology laboratory; DDJB, DNA data bank of Japan; PTM, post-translational modification

Opinion

Liberal arts education provides time and space for young researchers to explore moral, ethical, legal and technical challenges posed by the precision medicine initiative

The American Association for the Advancement of Science defines a liberal arts education that “produces persons who are open-minded and free from provincialism, dogma, preconception, and ideology; conscious of their opinions and judgments; reflective of their actions; and aware of their place in the social and natural worlds.¹ This definition articulates the potential opportunities for students trained in this tradition to make contributions in the goals of the Precision Medicine Initiative introduced by President Obama in 2015 to integrate many fields of study: Law; Ethics; Morality; Computer Science; Mathematics; Biology; and Chemistry, to deeply reflect on the human impacts of initiatives such as these. The curricula at Liberal Arts colleges provide students freedom, time and space to pursue independent studies and independent research projects to delve into cross-disciplinary questions. In our view, there are two primary ways in which investigators at Liberal arts colleges can contribute to the Precision Medicine Initiative. Firstly, skill sets in

computational sciences can be harnessed to develop connections between biochemical features of disease progression in humans to well characterized mechanisms described in pre-clinical animal model systems. Secondly, students from many different backgrounds are housed in a single institution and are required to take a broad range of courses providing investigators a unique opportunity to survey opinions and knowledge of these diverse set of students to better inform impacts and knowledge of large public initiatives. Liberal Arts setting provide a forum for undergraduate students to explore the technical, moral, ethical and legal aspects of the Precision Medicine Initiative.

Advanced open source computational tools allow for unprecedented analysis of publicly available data linking clinical outcomes to biochemical mechanisms to deliver goals of the precision medicine initiative

President Obama’s Precision Medicine Initiative launched in 2015 (<https://www.whitehouse.gov/the-press-office/2015/07/08/fact-sheet-new-patient-focused-commitments-advance-president%E2%80%99s-precision>) commits Federal agencies and private-sector groups to: Increase portability of health data; Make it easier for patients to access and share their own health information; Protect patient privacy, security and choice; and support new research platforms connecting researcher and participants as partners, with the mission to “articulate a set of core values and responsible strategies for building public trust and maximizing the benefits of a large national research cohort, while minimizing the risks inherent in large-scale data collection, analysis and sharing”. In response to this the FDA has created “precision FDA” research and development portal that allows

for testing, piloting, and validating existing and new bioinformatics approaches to next generation sequence processing (<https://precision.fda.gov/>) and enforcements of new standards for electronic submissions for the drug approval process (<http://www.fda.gov/ForIndustry/ElectronicSubmissionsGateway/ucm2005551.htm>). The rapid advance of software platforms such as Amazon Web Services, Azure and Google cloud enable housing of massive amounts of health data compliant with Health Insurance Portability and Accountability Act and Genetic Information Nondiscrimination Act. This data can be accessed either through public repositories (DDJB/EMBL/Genbank sequence databases, DDJB (Japan) and EMBL (Europe) and Genbank (USA) or through licensing agreements with companies like Symphony Health Solutions (<http://symphonyhealth.com/>) that houses petabytes of information garnered from thousands of patient, payer and provider sources. The standardized format of metadata and the development of numerous programming platforms that are non-relational, distributed, open-source and horizontally scalable such NoSQL (<http://nosql-database.org/>) provide for powerful methods to link data from many different sources such that genetic or biochemical signatures identified in patients can be attributed to potential basic mechanisms of transcription, translation, DNA modifications and protein modifications.

Opportunities for integrating clinical sciences and basic sciences have never been more promising. In this paradigm multiple databases can provide insights into mechanism of disease to clinical manifestation: Structural information following transcription is provided for by a database, RNA STRAND, that contains thousands of RNA secondary structures and information for potential binding sites for splicing proteins that could lead to the diversity of observed isoforms of spliced regulatory proteins (<http://www.masoft.ca/strand/>); Post-translational modifications (PTMs) have been cataloged in SysPTM containing data detailing more than 100,000 experimentally determined PTM sites on tens of thousands of proteins involving nearly 50 PTM types, curated from public resources including five databases and four web servers and more than one hundred peer-reviewed mass spectrometry papers; Protein annotations including Pfam domains, KEGG pathways, GO functional classification, and ortholog groups are integrated into the database. Four online tools have been developed and incorporated, including: PTMBlast, to compare a user's PTM dataset with PTM data in SysPTM; PTMPathway, to map PTM proteins to KEGG pathways; PTMPhylog, to discover potentially conserved PTM sites; and PTMCluster, to find clusters of multi-site modifications (<http://lifecenter.sgst.cn/SysPTM/index.jsp>). The hope is that it will yield new insights of the disease process and targeting patient specific proteins or pathways for treatment options. At Gustavus Adolphus College, the web-based tools have proved invaluable in pursuit of numerous Bioinformatic projects undertaken by students. Recent projects included: "Nano-Immunotherapy. Cancer Genes in Pre-Clinical Mouse Models" by Janessa Anderson; "Pharmacogenomics: Role of MYC in Leukemias" by Shelby Roles; "Genetics of Cancer" by Sara Kullberg; "Leukemogenesis: Mutations in cancer progression" by Sara Kullberg; "Cancer Mutations activating pathways." By Shelby Roles; "Leukemogenesis. Alternative splicing in cancer." by Shelby Roles; "Next Generation Sequencing in targeted therapies." by Rakim Solaiman. Earlier projects with former students led to peer reviewed articles in the British Journal of Haematology,² Expert Reviews of Anticancer Therapies³ and E Biomedicine.⁴ The liberal arts setting allowed the research mentor to discuss in detail both the scientific and humanitarian aspects of genomic in preclinical and clinical research. Students gained knowledge about how targeted therapies are discovered and developed for clinical use.

Surveying ethical and knowledge base of students with regard to modern genetic based research and therapies

Liberal Arts Colleges are uniquely situated to consider moral, ethical and legal issues surrounding issues in the post-genomic era and the adoption of precision medicine in healthcare. In order for progress to continue to be made, there must be a moral, ethical, and legal framework for research to be done, otherwise public opinion and politics could roadblock research such as when the Bush administration banned embryonic stem cell research funding in 2001.⁵ Specifically, lobbying committees who oppose embryonic stem cell research such as the National Right to Life and the US Conference of Catholic Bishops were able to exploit a lack of public knowledge about stem-cell research to achieve their goals.⁶ In WGS, the main ethical concern is privacy and unwanted information. For instance, how can an individual be protected from accidentally finding out that they have an untreatable disease or inadvertently deriving that information from the results of a blood-relative's genetic tests? In gene therapy, the two main dilemmas are 1) what diseases should qualify for gene therapy considering the risk and expense, and 2) how can minority populations (i.e. the deaf community) be protected from pressure to undergo gene therapy if it were made available to them? Lastly, for epigenetics, should a mother or father be required to quit a behavior (e.g. smoking) that could negatively modify their genome in a way that is heritable to their children?

Individuals with specialized knowledge are required for advancing ideas and cross-disciplinary concepts in this framework for the PMI; specifically, individuals who are educated in both the biological and ethical aspects of the technology are needed. Arguably, the only place where individuals are educated in this combination of topics occurs in large numbers is at undergraduate liberal arts institutions. This realization prompted us to investigate the genetic technology knowledge base of undergraduates at Gustavus Adolphus College. We hypothesized that there would be a strong familiarity of both new genetic technology as well as the policies surrounding it (Appendix 1). To measure this, we administered a 30 question survey to 90 students at Gustavus Adolphus College (Table 1) to measure familiarity with and acceptance of new genetic techniques (WGS, GT, and epigenetic therapy) and were surprised to find a low understanding of both the technology and policies (statements 1, 12, 22 show a low knowledge of WGS, GT, and epigenetics respectively, statements 8 and 9, measure knowledge of GINA, a law protecting discrimination based on genetic test results; Figure 1 & Table 2). What is even more troubling is that students are highly accepting of techniques that are not well understood either technologically or legally (statements 2, 13, and 27 show a high acceptance of WGS, GT, and epigenetics respectively; Figure 1 & Table 2). This is troubling because central to tens and thousands of people participating in large scale sequencing initiatives require informed consent. Surveys conducted at liberal arts institutions are capable of detecting lack of knowledge that could be utilized to develop more appropriate informed consent documents from both the researcher's end and the subject's end being one of the major goals of the Precision Medicine Initiative. Inadequate informed consent not only makes the researcher vulnerable to major liability but can also decrease the trust of the public in researchers in this field and reduce large scale participation of subjects in research trials.⁷⁻⁹

However, another interesting trend is that increasing education in biology was correlated with increased acceptance of these techniques (Figure 2). Because of this, we concluded that early introduction of bioinformatics technology into a science curriculum would be

a reasonable technique for increasing public acceptance of these techniques. We propose extensive training and use of higher level programming languages hosted by servers in the cloud environment to data mine and analyzes very large datasets. Coupled with gaining

this technical knowledge students need to participate in discussions exploring ethical, moral and legal ramifications of these genomic technologies especially related to informed consent issues to increase public trust and participation in large scale clinical trials.

Table 1 Description of survey respondents from the three classes at Gustavus Adolphus college

Class	Number of participants (males, females)	Class standing of participants	Percentage biology majors
January-term knitting class	24 (7,17)	First-year (14), Sophomore (8), Junior (1), and Senior (1)	0%
100-level biology class	43 (20, 23)	First-year (39) and sophomore (4)	63%
300-level genetics class	23 (6,17)	Junior (5) and senior (18)	100%

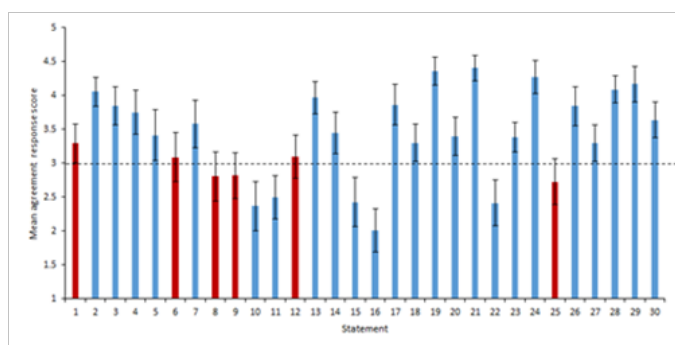


Figure 1 Mean agreement response score \pm 99.5% confidence interval for the 30 ethics of genetics survey questions (n=90). Red bars indicate that the average score did not differ significantly from the neutral score of 3 (represented by a dashed line), blue bars did differ significantly from 3.

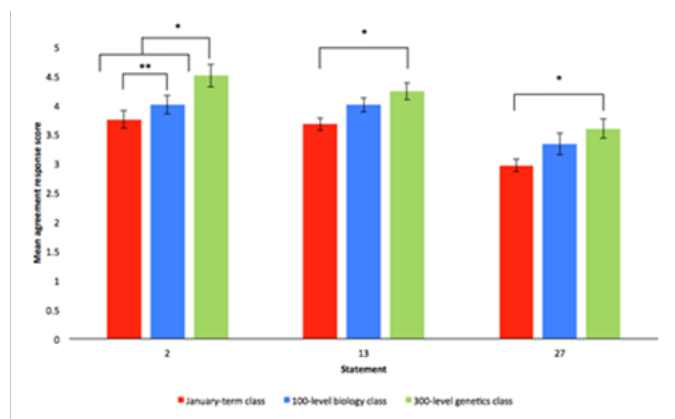


Figure 2 Mean agreement responses to three select questions from the genetic technique acceptance survey for three classes at Gustavus Adolphus College: a January-term class (n=24), a 100-level biology class (n=43), and a 300-level genetics class (n=23).

The error bars represent 95% confidence intervals. Increasing education correlates with increased acceptance of the genetic technique. *indicates significance ($p < 0.05$) and **i indicates borderline significance ($0.05 < p < 0.1$).

R and shiny programming language provide an accessible and integrated platform that can be housed on cloud resources for accessing and analyzing bioinformatic data with relevance to precision medicine initiative

However, a small budget and limited access to expensive bioinformatics equipment and personnel make it challenging for most liberal arts establishments to integrate these technologies into their curriculum. We propose two approaches to overcome these challenges.

First, open source tools (such as 1000 genomes and mutation taster) can be used to simulate and expose students to bioinformatic research techniques. Specifically, a professor could assign each student a made up gene variant with the premise that it was discovered via WGS. Students could then use these two tools to provide evidence indicating whether this variant is a polymorphism or mutation. A second approach would be to partner with core facilities to provide tours to students and show them what research in this field looks like. Hands on experience with this technology would be helpful in making these highly-conceptual processes less abstract to students. A partnership with a bioinformatics core would be easier to achieve for a liberal arts institution because the smaller class sizes would be easier to accommodate by the core facility.

In order to take advantage of these opportunities, the curriculum needs to create a broad platform of independently creative students in the Arts, Humanities and Sciences to work together in a collaborative and creative climate as a single innovation team with a common vision of innovation, deep conviction and persistence to problem solve an unmet need in biomedicine. This will require an agile and dynamic work environment that is enabled by software tools responsive to the needs of these types of teams. The Bioinformatics program at Gustavus Adolphus College developed in-house applications to enable exploration of issues in the clinical setting to gain a better understanding of systems aspects of biochemical regulation of gene expression in human diseases. The major goal of our research was to perform meta-analysis of published genomic databases to identify novel biochemical pathways in the leukemia genome that leads to early relapse in childhood cancers. Expansion of drug/radiation resistant clones is likely to involve co-ordination and action of hundreds of genes in numerous apoptotic and anti-apoptotic pathways. We are developing a database and visualization tools for B-Cell precursor acute lymphoblastic leukemia to explore genomic level expression of cancer genes to synergize efforts at the research level. The Bioinformatic Program at Gustavus Adolphus College worked in close collaboration with Dr. Uckun at Children's Hospital Los Angeles/ Kenneth Norris Jr. Comprehensive Cancer Center at USC from 2008 to 2015 and published 39 articles in peer reviewed journals developing novel therapeutic approaches against childhood leukemias. This highly productive effort was aided by the computational and bioinformatic web-resources for genomic data analysis. This effort led to the characterization of important intronic mutations in the CD22 receptor resulting in exon loss and activation of cellular proliferation mechanisms,¹⁰ activation of antiapoptotic transcription factor STAT3 serving as a substrate for SYK resulting in B-lineage leukemia/lymphoma,¹¹ discovery of CD19 ligand as a proapoptotic ligand¹² and the genetic signature for development of resistance in infant leukemias.¹³ We focused on a discovery-driven approach because currently, therapeutic advances in oncology are severely hampered by the high

rate of clinical stage failures demonstrated by 95% of drugs failing human clinical trials and success rate for translation of therapies from animal models to human trials is less than 10%. Our area of focus is to develop preclinical mouse models that better represent the disease in human subpopulations by directly incorporating the human Oncogene in a transgenic mouse and characterize the progression of the disease in the mouse and comparing the genetic mechanisms known to occur in the human cancers. The Bioinformatic tools enable us to compare human and mouse genomes to develop model systems to study disease progression. Presently there is an unmet need for highly targeted therapies for infant leukemias to reduce toxicity of chemo/radiation therapy and treat the rapid development of resistant clones of cancer cells. Data mining, bioinformatics, computational modeling and simulation have increasingly become integral to development of novel hypotheses in disease progression and potentially novel therapeutic innovations that target prominent receptors in leukemias (CD19 and CD22 receptors) to eradicate cancer cells.^{14–20} Our project attempted to link data mining efforts that identify genetic signatures in high risk patient sub-groups to data explorations methods developed using R made available in Bioconductor (<http://www.bioconductor.org/>) by writing interactive, exploration based statistical analysis platform provided by Shiny applications (<http://shiny.rstudio.com/>). The main data repository, Gene Expression Omnibus (GEO), is an Array- and sequence-based database that contains nearly 3500 datasets and more than one million samples. Bioconductor harbors more than 800 software packages specifically designed to analyze and explore array based data. Our collaboration identified novel high

risk patient molecular sub populations from GEO and implemented data exploration tools in Shiny to identify potential cell proliferative pathways for therapeutic exploitation. Through development of data visualization tools deployed in this software environment, we were able to employ interactive heatmaps, density plots, box-whisker plots, multivariate analysis of variance to identify potential cancer genes with HTML/Javascript code rendered in Shiny. Future work will utilize statistically powerful inferential and predictive algorithms to assess the role of identified cancer genes in different types of high risk cancer patients through packages provided by the Bioconductor project (<https://www.bioconductor.org/>). The success of our efforts will lead to better data analysis exploratory tools that can be customized to teaching Biology students the power of current Data Science tools and consulting with small research groups that are increasingly employing high throughput technologies in their labs. The application of Shiny and R in the Linux environment will also allow us to exploit cloud resources for handling large data sets whereby calculations need to be performed in parallel and requiring large working memory to perform the calculations. This is a very attractive proposition for developing teaching modules that introduce students to cloud computing such as Amazon Cloud EC2. Azure/Google cloud without requiring large budgets for acquiring in house computer hardware. This scientific knowledge coupled with well-versed understanding of moral, ethical and legal issues will produce broadly trained investigators in the biomedical setting with the capability to advocate for PMI and recruit both patients and collaborators to meet the goals for PMI.

Table 2 Mean agreement response score and conclusion of survey statements

Statement	Mean response score	Conclusion
2, 3	4.1, 3.8	Whole genome sequencing accepted for both newborns and adults.
4, 7, 10, 11	3.8, 3.6, 2.4, 2.5	Respondents interested in having genome sequenced but have conflicting feelings about discrimination and loss of privacy based on results.
13 and 14	4.0, 3.4	Gene therapy accepted for disease treatment in children and adults.
15 and 16	2.4, 2.0	Gene therapy not accepted for enhancements in children or adults.
17	3.9	Respondents trust doctors to inform them on gene therapy.
18–21	3.3, 4.4, 3.4, 4.4	Respondents interested in gene therapy for both life-threatening and non-life-threatening but chronic disease.
22	2.4	Respondents unfamiliar with epigenetics.
23	3.4	Respondents believe environmental factors can alter the epigenome.
24, 26, 27	4.3, 3.8, 3.3	Parents should protect children from negative epigenetic factors and epigenetic treatments should be allowed.
28–30	4.1, 4.2, 3.6	Respondents believe education affects their beliefs and that they have enough access to it.
1, 12	3.3, 3.1	Respondents neither familiar nor unfamiliar with WGS and gene therapy
6	3.1	Respondents neither fearful nor fearless of discrimination from employer
8, 9	2.8, 2.8	Respondents not familiar with laws protecting them from discrimination based on genetic test results
25	2.7	Respondents are neutral with regards to considering epigenetic consequences when making daily lifestyle decisions.

Acknowledgements

None.

Conflict of interest

The author declares no conflict of interest.

References

1. Project on Liberal Education and the Sciences. *The Liberal Art of Science: Agenda for Action*. Washington, DC: AAAC ISBN 0–87168–378–4. 1990.
2. Uckun FM, Qazi S, Ozer Z, et al. Inducing apoptosis in chemotherapy-resistant B-lineage acute lymphoblastic leukaemia cells by targeting HSPA5, a master regulator of the anti-apoptotic unfolded protein response signalling network. *Br J Haematol*. 2011;153(6):741–752.
3. Uckun FM, Pitt J, Qazi S. JAK3 pathway is constitutively active in B-lineage acute lymphoblastic leukemia. *Expert Rev Anticancer Ther*. 2011;11(1):37–48.
4. Uckun FM, Ma H, Ozer Z, et al. A previously unknown unique challenge for inhibitors of SYK ATP-binding site: Role of SYK as a cell cycle checkpoint regulator. *EBioMedicine*. 2014;1(1):16–28.

5. Park A, George W. Bush and the stem cell research funding ban. *Time Magazine*. 2012.
6. Wertz DC. Embryo and stem cell research in the United States: history and politics. *Trends Mol Med*. 2002;8(3):143–146.
7. Hayden EC. Informed consent: a broken contract. *Nature*. 2012;486(7403):312–314.
8. Allen C, Foulkes WD. Qualitative thematic analysis of consent forms used in cancer genome sequencing. *BMC Med Ethics*. 2011;12:14.
9. Dresser R. Research ethics. Aligning regulations and ethics in human research. *Science*. 2012;337(6094):527–528.
10. Uckun FM, Goodman P, Ma H, et al. CD22 EXON 12 deletion as a pathogenic mechanism of human B–precursor leukemia. *Proc Natl Acad Sci U S A*. 2010;107(39):16852–16857.
11. Uckun FM, Qazi S, Ma H, et al. STAT3 is a substrate of SYK tyrosine kinase in B–lineage leukemia/lymphoma cells exposed to oxidative stress. *Proc Natl Acad Sci U S A*. 2010;107(7):2902–2907.
12. Uckun FM, Sun L, Qazi S, et al. Recombinant human CD19–ligand protein as a potent anti–leukaemic agent. *Br J Haematol*. 2011;153(1):15–23.
13. Qazi S, Uckun FM. Gene expression profiles of infant acute lymphoblastic leukaemia and its prognostically distinct subsets. *Br J Haematol*. 2010;149(6):865–873.
14. Uckun FM, Mitchell LG, Qazi S, et al. Development of Polypeptide–based Nanoparticles for Non–viral Delivery of CD22 RNA Trans–splicing Molecule as a New Precision Medicine Candidate Against B–lineage ALL. *EBioMedicine*. 2015;2(7):649–659.
15. Uckun FM, Ma H, Cheng J, et al. CD22ΔE12 as a molecular target for RNAi therapy. *Br J Haematol*. 2015;169(3):401–414.
16. Uckun FM, Myers DE, Cheng J, et al. Liposomal Nanoparticles of a Spleen Tyrosine Kinase P–Site Inhibitor Amplify the Potency of Low Dose Total Body Irradiation Against Aggressive B–Precursor Leukemia and Yield Superior Survival Outcomes in Mice. *EBioMedicine*. 2015;2(6):554–562.
17. Uckun FM, Myers DE, Ma H, et al. Low Dose Total Body Irradiation Combined With Recombinant CD19–Ligand × Soluble TRAIL Fusion Protein is Highly Effective Against Radiation–Resistant B–Precursor Acute Lymphoblastic Leukemia in Mice. *EBioMedicine*. 2015;2(4):306–316.
18. Uckun FM, Myers DE, Qazi S, et al. Recombinant human CD19L–sTRAIL effectively targets B cell precursor acute lymphoblastic leukemia. *J Clin Invest*. 2015;125(3):1006–1018.
19. Uckun FM, Qazi S, Ma H, et al. CD22ΔE12 as a molecular target for corrective repair using RNA trans–splicing:anti–leukemic activity of a rationally designed RNA trans–splicing molecule. *Integr Biol (Camb)*. 2015;7(2):237–249.
20. Uckun FM, Qazi S, Cheng J. Targeting leukemic stem cells with multifunctional bioactive polypeptide nanoparticles. *Future Oncol*. 2015;11(8):1149–1152.