Quantum computing for clinical research

Investigator Team:
Shan Shan (SCIENCE), Jørgen Ellegaard Andersen (SCIENCE),
Maja Sofie Thiele (HEALTH), Maria Kjærgaard (HEALTH) Aleksander Krag (HEALTH)
Moustapha Kassem (HEALTH) and Maria Timofeeva (HEALTH)

March 29, 2022

1 Introduction

Big data in clinical research. The exponential growth of healthcare data – such as disease registries, electronic health records (EHRs), results of medical examinations and tests – pointed to an overwhelming need for developing methods that address the challenges of analyzing and interpreting big data to build healthier societies. Technical challenges include a range of issues from data collection, visualization of high-dimensional data, simulation of clinical trials, interpretability, feature selection, robustness, prediction, optimization, uncertainty quantification, model validation and evaluation. With recent advances in machine learning, artificial intelligence, deep learning, and high-performance cloud computing, significant progress has been made in data-based clinical research to accelerate diagnoses, optimize pricing, enhance patient experiences, and improve healthcare practitioner work lives. However, given the rapid pace of data collected and generated from individuals, devices and systems, the need for greater computing power is stretching the capabilities of classical computing systems.

The promise of quantum computers. Quantum computers can exponentially advance computational power and solve problems that cannot be tackled by classical machines. In health and biomedical research, quantum computers have the potential to enable numerous computation-intensive applications, such as supersonic drug design, in silico clinical trials simulation, medical imaging analysis. However, recent efforts in the field have mainly focused on the problems of protein folding [6][19] and protein design [15] using either a *quantum annealing approach* [10] and the *quantum gate approach* [1]. The quantum annealing approach recasts the research question as a *Quadratic Unconstrained Binary Optimization* (QUBO) program and solves the reformulated problem on an abiadiatic quantum device (e.g., The Advantage quantum computer developed by D-Wave). The quantum gate approach maps the classical algorithm to its corresponding quantum circuit and runs the circuit on a universal quantum computer, of which we are starting to see the first few prototypes in existence now and the technological development is expected to strongly accelerate over the coming years.

Advantage of Gaussian Boson Samplers. There has been little work that addresses other aspects of the above-mentioned challenges in healthcare research in the era of big data, nor using other emerging quantum technologies, such as *Gaussian Boson Samplers* (GBS) [9]. GBS is a photonic quantum model that has been demonstrated with quantum advantage on a special sampling task. In 2021, Zhong et al. [27] reported a 144-mode GBS that yields a sampling rate approximately 10^{24} times faster than using brute-force simulation on classical supercomputers. Compared to other leading quantum efforts, the biggest advantage of GBS is that GBS can work in room temperature[1], making it a promising candidate for wide use in the future.

The objective of this proposal is to develop machine learning methods in clinical research based on a new quantum technology – Gaussian boson samplers (GBS). The significance of this proposal is that enabling usage of GBS in healthcare would help exploit the near-term quantum technologies to their full potential, and opens a vast and uncharted space of quantum methods in health and clinical research.

[1]with the possible exception for the photon counters, some of which needs cooling to around 4K.

2 Specific Research Aims

As a first concrete target, we propose to explore use cases of GBS with the following research aims (A):

A1: Optimize pathways for accurate referral of fatty liver disease patients. We will develop a new GBS-based classification algorithm that identify patients with advanced liver fibrosis from three non-invasive bio-markers using simple blood tests and ultrasound-based screening tools. The new classification method will exploit the graph-theoretical framework for robust fitting as we have developed in [22], and will be suitable to deal with noisy data – for example, due to measurement error or missing tests. The new method can potentially improve and accelerate the diagnoses of advanced liver fibrosis using non-invasive tests.

A2: Improve the clinical efficacy of transplanted human bone marrow stromal cells. We will develop a new quantum-enhanced feature selection method to identify a subset from donor- and cell-related features for the ability of stem cells to form bone in vitro. We will first encode the information about the importance and correlation of features into a weighted graph, and then use GBS to find important features through clustering on the weighted graph. The proposed new feature selection technique could enhance classification accuracy and offer greater interpretability for the classification model.

A3: Enhance the prediction accuracy for colorectal cancer risk using genetic, biochemical and clinical data from UK Biobank. We will use the method proposed in A2 to identify key features in genetic, biochemical and clinical data to improve the prediction accuracy for colorectal cancer risk. We propose a new hybrid framework that enables the quantum-enhanced feature selection method of A2 to be implementable on a super large graph. The framework consists of first using a classical trimming algorithm to filter out inactive vertices, and then using GBS to cluster the remaining active vertices. The proposed technique enables application of GBS algorithms to massive graphs stemmed from real problems that often have millions of vertices and hundreds of millions of edges, and further improve the accuracy for colorectal cancer risk prediction.

Interdisciplinary aspect. The research aims (A1-3) can only be achieved through an interdisciplinary study that combines quantum computing, statistics, machine learning, health and biomedical research in order to explore the full potential of quantum technologies for real-world health challenges. The proposed research will build on joint efforts of the following personnel.

- Professor Jørgen Ellegaard Andersen (DIAS Chair of Quantum Mathematics) and Post-doc Shan Shan from Centre for Quantum Mathematics, whose expertise are in photonic quantum computing technologies and toplogical quantum field theory.
- Professor Aleksander Krag (DIAS Chair of Health Sciences), Professor Maja Sofie Thiele and PhD student Maria Kjærgaar from the Center of Liver Research.
- Professor Moustapha Kassem (DIAS Chair of Health Sciences) from KI, Endocrinology.
- Professor Maria Timofeeva (DIAS Fellow of Health Sciences) from the Department of Public Health.

Potential impact to science and society. One of the biggest challenges in the near-term quantum computing technologies is to identify specific problems of practical interest for which these devices can prove advantageous. The success of the new quantum algorithms developed in the proposal will significantly enlarge the applicability of the current and emergent quantum computing technologies. Healthcare and medicine are key fields for future high-impact quantum application because of their large and increasing demand for computing power. A quantum workforce within these fields will bring groundbreaking change to our everyday life. In addition to the three problems – diagnosing advanced liver fibrosis, identifying key features for osteoblastic differentiation in hBMSCs, and predicting colorectal cancer risk – as we proposed to study above, our methods also enjoy applications other problems. Density estimation, feature selection and classification are fundamental tasks in statistical machine learning. A quantum solution to these problem would potentially help address other important issues in society.

Relevancy to DIAS. The proposed research is truly interdisciplinary, because it tackles a challenging problem – whose solutions are beyond the scope of a single discipline – through integrating information, data, techniques, tools, perspectives, concepts from both clinical research and quantum computing. Thus, this proposal aligns with the mission of DIAS that "brings together outstanding researchers together in an interdisciplinary center for fundamental research and intellectual fora". While the proposed research focuses on applications in clinical research, such as liver fibrosis, bone formation, and colorectal cancer, the quantum-based machine learning techniques we will be developing in this research have a wide application in all data-intensive fields. This project will be a first step towards promoting a cross-disciplinary approach to quantum computing which brings together experts in quantum mathematics, engineering, computer science, psychology, history, and other research fields at DIAS. We will discuss this aspect in more details

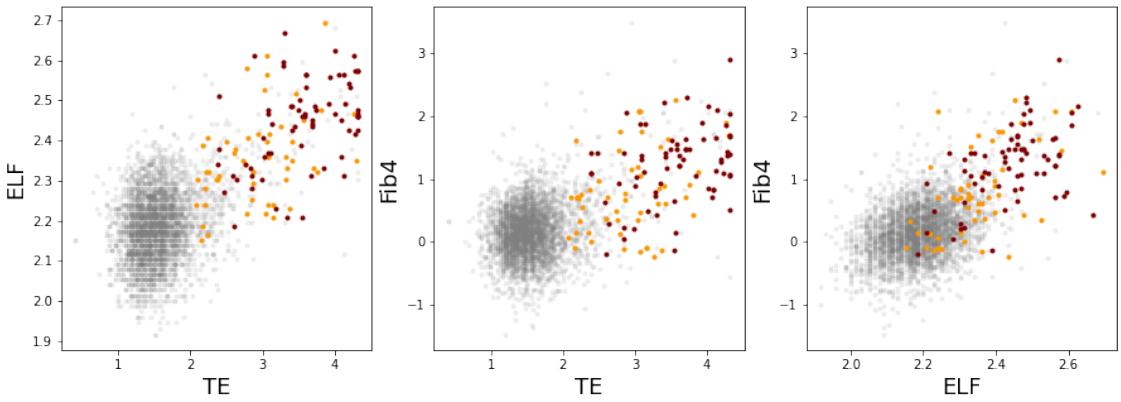

Figure 1: Log-log scatter plot of three non-invasive bio-markers (ELF, TE, Fib4) for diagnoses of liver fibrosis. Colored points represent patients with a Kleiner score $F = 3$ (yellow) or $F = 4$ (red). The gray points represent patients without a Kleiner score or whose Kleiner score is less than 2 ($F = 0, 1$ or 2).

through the planned activities in Section 4.

3 Research Strategy

We now describe in detail the research strategy for each research aim below.

A1: Optimize pathways for accurate referral of fatty liver disease patients.

Background. Liver fibrosis predicts liver-related and cardiovascular outcomes in chronic liver disease patients. Traditionally, evaluating liver fibrosis relies on histopathological examination of a liver specimen obtained by percutaneous biopsy [3]. However, the procedure for taking the liver biopsy is invasive and is poorly accepted by patients due to pain that results from the surgical operation. In recent years, many non-invasive tools have been developed to make painless, rapid and easy-to-perform assessment for liver fibrosis. The most commonly used non-invasive methods include: (1) Enhanced Liver Fibrosis (ELF) [17], a simple blood test to assess risk of fibrosis progression and liver related events; (2) Transient Elastrography (TE) [20], an ultrasound-based screening tool for liver fibrosis; (3) Fib4 score [24], a combination of patient age, platelet count, AST and ALT-cell tests done in a blood test. Although statistical correlation has been established between liver fibrosis and these non-invasive test scores [4], it remains an action of art to determine the precise cut-off values when using the non-invasive methods for diagnosis of significant fibrosis.

Preliminary data. The data that will be used in this proposal consists initially of 5266 patients, in which 5235 of them underwent TE, 3848 underwent ELF and 4524 received Fib-4 score. Among all the patients, 533 also received liver biopsy and obtained a Kleiner score [11], which is evaluated based on 14 histological features on the biopsies such as steatosis, lobular inflammation, hepatocellular ballooning and fibrosis. Severe fibrosis is indicated with a Kleiner score greater or equal to 3 ($F \geqslant 3$). To address this issue, we will develop a robust classification algorithm to identify severe fibrosis patients using their ELF, TE and Fib4 scores.

Research design. The main idea of the classification algorithm is to fit to data a multivariate normal distribution. Points that lie further away from the center are considered as outliers and therefore are classified as patients with severe liver fibrosis $F \geqslant 3$. Figure 1 illustrates the distribution of the log-transformed ELF, TE and Fib4 data. Note that the majority of data for patients with less severe liver fibrosis ($F \leqslant 2$) or those who do not receive a Kleiner score are concentrated approximately within an ellipsoid near small values of ELF, TE and Fib4. Patients with severe liver fibrosis ($F \geqslant 3$), marked with yellow or red points, lie mostly outside of the ellipsoid. We will model this ellipsoid by the *confidence ellipsoid* induced from the fitted normal distribution. The confidence ellipsoid of n standard deviation is given by

$$\text{CE}_n = \{x \in \mathbb{R}^3 \mid MD(x) = n\} \tag{1}$$

where the Mahalanobis distance $MD(x)$ is given by

$$MD(x) = \sqrt{(x-\mu)^\intercal \Sigma^{-1} (x-\mu)} \tag{2}$$

with μ and Σ denote the mean vector and covariance matrix of the multivariate normal distribution respectively.

Estimation of μ and Σ of the multivariate normal distribution requires a robust scheme because of the presence of gross outliers in the given data and the estimates μ and Σ are extremely sensitive to noise. Towards this, we will first use GBS to remove the gross outliers in data. In [22], we have developed a robust method for fitting noisy data to a parameterized model using GBS. Here, the parameterized model is the multivariate normal model, and the optimal parameters we are seeking are μ and Σ. However, the lack of a natural fitness score or residual function makes it difficult to directly apply our previous method. Instead, we will preform the outlier removal procedure progressively and work with one dimension at a time. From our previous paper, we know how to robustly fit lines and planes to noisy data as well as how to identify outliers from these models. In this problem, we will first find the best fitting plane to the three non-invasive bio-markers (ELF, TE, Fib4), and remove the outliers using the GBS method. Then, project all points to the fitted plane, and find the best fitting line of the projected points. Again, we will be using the GBS method and its outlier removal procedure. Finally, we project the 2D points onto the fitted line, and further identify the outlier points with the five-summary statistics. After the outliers are all removed, we will then use the sample mean and sample covariance to estimate μ and Σ, that is

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i, \;\; \Sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^{\mathsf{T}}. \tag{3}$$

Innovation. The proposed algorithm builds on the state-of-the-art quantum robust fitting framework developed in [22]. The novelty in this new algorithm is to extend the existing framework to robustly estimate mean and covariance, which is one of the fundamental tasks in robust statistics. Furthermore, the proposal will lead us to a new medical application using GBS, that will optimize referrals of fatty liver disease patients with non-invasive tools.

A2: Improve the clinical efficacy of transplanted human bone marrow stromal cells.

Background. Bone marrow stromal cells are a heterogeneous population of cells that contain a subset of skeletal stem cells that form cartilage, bone, and support hematopoiesis and formation of marrow adipocytes [2]. Transplantation of human bone marrow stromal cells (hBMSCs) is a promising therapy to enhance tissue regeneration following injury [26] [23] [16]. As hBMSCs are being tested in a wide variety of early phase clinical trials, efficacy of the therapy with respect to bone regeneration varies among trials [7] [16]. Previous studies have worked towards identifying a set of clinical signature that will predict hBMSCs with optimal osteoblastic differentiation capacity [12]. Kowal et al. performed a multivariate linear regression analysis with step-wise backward selection collectively on the clinical signatures, and identified a set of collective variables that predicts > 50% of variation in osteoblast and adipocyte differentiation outcome. In this project, we will extend their work to build a predictive model for the osteoblast and adipocyte differentiation outcome by developing a quantum-enhanced method.

Preliminary data. The data that we will be using in this project come from the study in [12], which consists of 58 patients undergoing surgery for bone fracture. The clinical profile of each patients include 96 donor- and cell-related characteristics. For a detailed description of the clinical characters, see [12].

Research design. Our prediction model is built from two steps. The first step is to identify a subset of clinical characteristics with significant impact on the osteoblastic differentiation capacity. The second step is to train a predictive model from the selected feature characteristics. We need the feature selection step (step 1) because the feature space in our data is of high dimensionality ($\geqslant 96$), making it difficult to deliver accurate results for almost all classification algorithms. Instead of linear regression and backward selection as done in [12], we will use a clustering approach to identify relevant features. The application of cluster analysis has been demonstrated to be more effective than traditional feature selection algorithms [18] [5]. Our contribution is a novel clustering algorithm using the new quantum technology GBS, which we briefly describe as follows.

The basic idea is to construct a weighted graph where each vertex represents a clinical character and edges encode correlation between two clinical characters. More precisely, the edge weight between clinical characters v_1, v_2 is assigned to be their *Pearson's Correlation Coefficient*,

$$\rho = \frac{cov(v_1, v_2)}{\sigma_{v_1}\sigma_{v_2}}, \tag{4}$$

where $cov(v_1, v_2)$ denotes the covariance between the two characters and σ_{v_i} denotes the variance of v_i. We further assign the vertex weight of v_i to be the Pearson's Correlation Coefficient between v_i and osteoblastic differentiation capacity. Once the graph is constructed, we will use GBS to sample several large and heavy cliques – fully connected

subgraph with large vertex weights. Then we form clusters on the graph by joining cliques which largely overlap. Last, we select from each cluster the most representative feature that is strongly related to osteoblastic differentiation to form a subset of features.

Innovation. From the machine learning perspective, the proposed method in this research aim is a novel application of GBS to feature selection, which is one of the most critical tasks of big data analysis and could be further applied to a wide range of applications. From the medical perspective, the proposed method, if successfully implemented, could offer the clinicians and healthcare researchers more insights about the hBMSCs mechanism, and improve the success of the hBMSCs transplantation therapy.

A3: Enhance the prediction accuracy for colorectal cancer risk using genetic, biochemical and clinical data from UK Biobank.

Background. Colorectal cancer (CRC) is one of the most common types of cancer worldwide and the third leading cause of cancer death in 2020 [25]. Given that the heritability of CRC has been estimated to be around 16% to 35% and the sibling recurrence risk ratio is about 2.0 [8], genetic susceptibility could contribute greatly to CRC risk. Previous studies estimated that about 40% of the variation in CRC risk can be attributed to genetic factors and identified more than 200 genetic variants associated with the colorectal cancer risk [13]. While common genetic variants are correlated with CRC risk, using them to predict the risk showed very limited discrimination ability, even after inclusion phenotypic factors such as sex, age, family history, BMI and diet [14][21]. One possible explanation to the poor performance of the current prediction models is that these models overlooked key predictors such as clinical history (e.g., whether the individual has inflammatory bowel disease or diabetes) and biochemical data. To improve the prediction accuracy for colorectal cancer risk, we propose to use the wealth of the clinical, biochemical as well as genetic data generated by the UK Biobank to find new predictors for colorectal cancer risk.

Preliminary data. The UK Biobank (UKBB) is a large-scale biomedical database which contains detailed genetic, health and behavioral information from half a million UK participants aged between 40-69 at recruitment. All UKBB participants were genotyped using Affymetrix arrays and imputed using the Haplotype Reference Consortium and UK10K resources, which increased initial number of genotype variants to ca. 96 million variants. Information on 30 biochemistry biomarkers and clinical history are also available for the majority of participants. The data that will be used in this project consists of 6360 CRC cases and 25440 population-based control individuals. The access to UKBB is available through the project "Omics markers of age-related diseases and mortality" (ID75436, PI. M.Timofeeva).

Research design. The overall pipeline for predicting CRC risk resembles the model we proposed in A2, which consists of two steps – (1) feature selection by clustering on the correlation graph with GBS and (2) train the predictive model on the selected features. However, A3 represents a new challenge, that is the capacity (e.g., number of available qubits) of today's quantum hardware is still modest compared to the computational demand in real applications. The graph constructed in predicting CRC risk consists of more than 7 million nodes (– there are more than 7 million features combining the non-monomorphic genetic variants, clinical history and biochemistry markers), but the near-term GBS devices are only suitable for graphs of a few hundred nodes. There is a limitation on the input problem size that can fit on the current quantum hardware. To address this limitation, we propose a hybrid framework that utilizes a classical trimming algorithm to filter out inactive vertices in the graph, and then use GBS to cluster the remaining active vertices. The trimming algorithm finds the most active vertices based on a locality statistic and builds a smaller graph over active vertices. Then we assign the most active vertices into clusters through the algorithm outlined in A2.

Innovation. The innovation of this proposal is two fold. First, the proposal makes it possible to use GBS to solve clustering problems on massive graphs with sizes exceeding the number of available qubits on the near term quantum hardware. Second, compared to the existing methods for predicting CRC risks which used only a few top genetic variants, our proposal searches from a more comprehensive list of feature characteristics including genetic data, clinical records and biochemical biomarkers. If successfully implemented, our method can significantly improve the accuracy of CRC risk classification and therefore optimize prevention programmes for CRC in the population, for example through targeting screening, and also preventative interventions.

4 Planned Activities

To ensure the achievement of the outlined research aims (A1-3) and to enhance DIAS mission on promoting multi-disciplinary research, we plan for the following activities at DIAS during the project cycle.

Team meetings. We will hold weekly team meetings at DIAS on Wednesdays to increase project outcomes. During the team meetings, we will keep track of high-level research goals, identify development needs and address barriers to the research goals. The meeting provides an opportunity to foster collaboration and integrate knowledge from all members of the team.

DIAS day. Personnel funded by the project are committed to spend each Wednesday at DIAS to increase interaction with the larger DIAS community. DIAS, being the interdisciplinary elite research center at SDU, is home to many world-leading scholars in science, humanities, health. We are hoping to promote synergies between these diverse teams at DIAS by organizing informal and formal events that gather researchers and students. As a first step towards this goal, DIAS day will then facilitate informal discussions and collaborations from our team to the larger DIAS community.

DIAS workshop. We are planning to organize a small workshop on quantum computing for clinical research at DIAS. The workshop will gather research leaders from the Faculty of the Health Sciences (SUND) at SDU to encourage collaboration and further work within SDU. We will of course also invite all other DIAS chairs and fellows to participate in this workshop, since the application of quantum computing technology has a much wider scope than the medical discipline. We hope this activity will bring together a group of researchers at DIAS and SDU to collaborate using the new quantum computing technology.

DIAS lecture series. We will organize a series of lectures on quantum computing and its application to clinical research at DIAS during the project cycle. The lecture series will introduce the basics of quantum computing and its application to clinical research. We will cover fundamental quantum building blocks, photonic quantum computing, quantum algorithms. While we study applications in clinical research, we focus on the underlying tools, which will make the course appealing to a broad audience.

5 Possible funding for the project going forward

A grant from DIAS would allow us, as indicated above, to establish a pilot project and most likely produce the first few papers, which demonstrate a proof of concept of our interdisciplinary approach. This would give us the perfect offset for applying to other foundations for an expansion of this project, in particular some of the Novo Nordisk Foundations programs that target life science and health data analysis, but also a number of programs at OUH's forsknings og innovationspuljer and possibly also the Independent Research Fund Denmark (DFF).

6 Budget

Budget	2022							
	Total		Applied		Co-fin.		Other	
LØN	Md	DKK	Md	DKK	Md	DKK	Md	DKK
Shan Shan (5/2022 – 10/2022)	6,0	301.000	6,0	301.000				
LØN I ALT	6,0	301.000	6,0	301.000	0,0	0	0,0	0
DRIFT		DKK		DKK		DKK		DKK
Lecture series		60.000		60.000				
Workshop		24.000		24.000				
DRIFT I ALT		84.000		84.000		0		0
OMKOSTNINGER I ALT	6,0	385.000	6,0	385.000	0,0	0	0,0	0
Overhead								
BUDGET I ALT	6,0	385.000	6,0	385.000	0,0	0	0,0	0

Table 1: Budget summary.

We budget the following items for this project. The budget summary is also reported in Table 1.

Personnel. Shan Shan, post-doc, six months salary (01/05/2022 - 31/10/2022). Dr. Shan's expertise is in

high-dimensional data analysis and quantum machine learning. She has a long working relationship with Professor Jørgen Ellegaard Andersen. Dr. Shan will contribute to the design of the quantum algorithms and the implementation of the data analysis in this project.

Support for the workshop. To cover workshop materials and catering (lunch and coffee breaks) we budget the program cost for the DIAS workshop at a total of DKK 24.000,- (40 participants for 4 days for an average of 150 per person per day).

Support for the lecture series. The program cost for the lecture series is budgeted at a total of DKK 60.000,-. This will cover 6 international speakers (a mix of European and overseas visitors) of an average of DKK 10.000 for each visit (travel, accommodation and meals)

References

[1] Adriano Barenco et al. "Elementary gates for quantum computation". In: *Physical review A* 52.5 (1995), p. 3457.

[2] Paolo Bianco et al. ""Mesenchymal" stem cells in human bone marrow (skeletal stem cells): a critical discussion of their nature, identity, and significance in incurable skeletal disease". In: *Human gene therapy* 21.9 (2010), pp. 1057–1066.

[3] Arturo A Bravo, Sunil G Sheth, and Sanjiv Chopra. "Liver biopsy". In: *New England Journal of Medicine* 344.7 (2001), pp. 495–500.

[4] Laurent Castera, Xavier Forns, and Alfredo Alberti. "Non-invasive evaluation of liver fibrosis using transient elastography". In: *Journal of hepatology* 48.5 (2008), pp. 835–847.

[5] Inderjit S Dhillon, Subramanyam Mallela, and Rahul Kumar. "A divisive information theoretic feature clustering algorithm for text classification". In: *The Journal of machine learning research* 3 (2003), pp. 1265–1287.

[6] Mark Fingerhuth, Tomáš Babej, et al. "A quantum alternating operator ansatz with hard and soft constraints for lattice protein folding". In: *arXiv preprint arXiv:1810.13411* (2018).

[7] Enrique Gómez-Barrena et al. "Bone regeneration: stem cell therapies and clinical studies in orthopaedics and traumatology". In: *Journal of cellular and molecular medicine* 15.6 (2011), pp. 1266–1286.

[8] Rebecca E Graff et al. "Familial risk and heritability of colorectal cancer in the nordic twin study of cancer". In: *Clinical Gastroenterology and Hepatology* 15.8 (2017), pp. 1256–1264.

[9] Craig S Hamilton et al. "Gaussian boson sampling". In: *Physical review letters* 119.17 (2017), p. 170501.

[10] Mark W Johnson et al. "Quantum annealing with manufactured spins". In: *Nature* 473.7346 (2011), pp. 194–198.

[11] David E Kleiner et al. "Design and validation of a histological scoring system for nonalcoholic fatty liver disease". In: *Hepatology* 41.6 (2005), pp. 1313–1321.

[12] Justyna Magdalena Kowal et al. "Identification of a clinical signature predictive of differentiation fate of human bone marrow stromal cells". In: *Stem cell research & therapy* 12.1 (2021), pp. 1–15.

[13] Philip J Law et al. "Association analyses identify 31 new risk loci for colorectal cancer susceptibility". In: *Nature communications* 10.1 (2019), pp. 1–15.

[14] Xue Li et al. "Prediction of colorectal cancer risk based on profiling with common genetic variants". In: *International Journal of Cancer* 147.12 (2020), pp. 3431–3437.

[15] Vikram Khipple Mulligan et al. "Designing peptides on a quantum computer". In: *bioRxiv* (2020), p. 752485.

[16] Regis J O'Keefe et al. "American Society for Bone and Mineral Research-Orthopaedic Research Society Joint Task Force Report on Cell-Based Therapies". In: *Journal of Bone and Mineral Research* 35.1 (2020), pp. 3–17.

[17] J Parkes et al. "Enhanced Liver Fibrosis (ELF) test accurately identifies liver fibrosis in patients with chronic hepatitis C". In: *Journal of viral hepatitis* 18.1 (2011), pp. 23–31.

[18] Fernando Pereira, Naftali Tishby, and Lillian Lee. "Distributional clustering of English words". In: *arXiv preprint cmp-lg/9408011* (1994).

[19] Anton Robert et al. "Resource-efficient quantum algorithm for protein folding". In: *npj Quantum Information* 7.1 (2021), pp. 1–5.

[20] Laurent Sandrin et al. “Transient elastography: a new noninvasive method for assessment of hepatic fibrosis”. In: *Ultrasound in medicine & biology* 29.12 (2003), pp. 1705–1713.

[21] Catherine L Saunders et al. “External validation of risk prediction models incorporating common genetic variants for incident colorectal cancer using UK Biobank”. In: *Cancer Prevention Research* 13.6 (2020), pp. 509–520.

[22] Shan Shan et al. “Robust fitting with Gaussian Boson Sampling”. In: *Under review at CVPR* (2021).

[23] Tiziana Squillaro, Gianfranco Peluso, and Umberto Galderisi. “Clinical trials with mesenchymal stem cells: an update”. In: *Cell transplantation* 25.5 (2016), pp. 829–848.

[24] RK Sterling et al. “S Sulkowski M, Torriani FJ, Dieterich DT, Thomas DL, Messinger D, Nelson M; APRICOT Clinical Investigators. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection”. In: *Hepatology* 43.6 (2006), pp. 1317–25.

[25] Hyuna Sung et al. “Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. In: *CA: a cancer journal for clinicians* 71.3 (2021), pp. 209–249.

[26] Walid Zaher et al. “An update of human mesenchymal stem cell biology and their clinical uses”. In: *Archives of toxicology* 88.5 (2014), pp. 1069–1082.

[27] Han-Sen Zhong et al. “Phase-programmable gaussian boson sampling using stimulated squeezed light”. In: *arXiv preprint arXiv:2106.15534* (2021).