School of Biological Sciences

ASSESSMENT COVER SHEET AND TEMPLATE

Section A – to be completed by the student

Student Number	**2228202**		
Programme	MSc Bioinformatics		
Unit Name	**Bioinformatics Project Planning**	Unit Code:	BIOLM0028
Assessment name	**Literature review**		
Word Count	3969		
Do you give permission for you work to be used anonymously in examples given to students in the future? (type Yes)			

By submitting this assignment cover sheet, I confirm that I understand and agree with the following statements:

- 'I have not committed plagiarism, cheated or otherwise committed academic misconduct as defined in the University's Assessment Regulations (available at https://www.bristol.ac.uk/media-library/sites/academic-quality/documents/taught-code/annexes/university-examination-regulations.pdf)
- 'I have not submitted this piece, in part or in its entirety, for assessment in another unit assignment (including at other institutions) as outlined in section 4 of the University's Assessment regulations (available at https://www.bristol.ac.uk/media-library/sites/academic-quality/documents/taught-code/annexes/university-examination-regulations.pdf)
- I understand that this piece will be scrutinised by anti-plagiarism software and that I may incur penalties if I am found to have committed plagiarism, as outlined in sections 3 of the University's Examination Regulations (available at https://www.bristol.ac.uk/media-library/sites/academic-quality/documents/taught-code/annexes/university-examination-regulations.pdf)

Title Page

The computational exploration of the cancer genome by liquid biopsies and cell free DNA

Glen Roarke

Abstract

Liquid biopsies (LBs) are emerging as a non-invasive complementary tool to explore the cancer genome alongside tissue biopsies. There are various LB biomarkers within precision oncology capable of providing genetic and epigenetic insight. Next generation sequencing (NGS) provides the foundation for LB clinical utility by analysing patient samples at high and low coverage, however computational bottlenecks exist.

Cell free DNA (cfDNA) is one biomarker used to detect circulating tumour DNA (ctDNA) to a high specificity and sensitivity, correlating with tissue biopsy analysis. cfDNA can provide therapeutic effectiveness in adjusting treatments, detecting minimal residual disease and longitudinal treatment response monitoring. ctDNA occurs at low quantities in early stage cancers but cfDNA and its epigenetic fingerprint show potential for early cancer detection by indicating chromatin accessibility, nucleosome positioning and DNA methylation. Various novel bioinformatics approaches differ in performance and require investigation.

Currently gene mutation panels are clinically more accessible than WGS approaches limiting epigenetic cancer profiling. Research into nucleosomal profiling software such as Griffin is required to produce clinically viable epigenetic assays. Generally, the genomic research field has ancestral limitations due to European bias which must be resolved to ensure equitable genomic assay performance in non-European patients.

Keywords

Liquid Biopsies, cell free DNA, circulating tumour DNA, biomarkers, precision oncology, nucleosome.

Challenging conventional cancer diagnosis using liquid biopsies

Liquid biopsies (LBs) have emerged as a promising tool within precision medicine, providing insight into the different aspects of disease such as diagnosis, progression and early detection (Heidrich et al., 2021). LBs can be undertaken on a range of different fluid types such as blood plasma, urine, cerebrospinal fluid and saliva, which can be utilised for different disease phenotypes (Otandault et al., 2019). The aim is to obtain normal or disease signals from the fluid sample via biological molecules called biomarkers. Consequently, there is extensive research ongoing into optimising techniques of blood collection, DNA extraction and isolation of biomarkers to produce clinical assays. Next, these assays need to pass regulatory approval, and undergo a cost analysis before being incorporated into clinical practice (Ignatiadis et al., 2021). It is common for whole genome sequencing (WGS) or whole exome sequencing (WES) to be used to generate targeted gene mutation panels for clinical diagnostics.

Precision oncology is one field where LBs have shown excellent potential to compliment clinical treatment in prevalent tumours such as breast (BC), colorectal (CRC) and lung cancer (LC) (Coombes et al., 2019;Zeinali et al., 2020), but also more rarer forms of prostate cancers (van der Toom et al., 2019). There are advantages in using LBs as an alternative to tissue biopsies, which are the preferred standard of cancer diagnosis. First, tissue biopsies can fail to identify certain genetic alterations within tumours due to intra-tumour heterogeneity or between different tumours due to inter-tumour heterogeneity (Gerlinger et al., 2012). More recent single-cell sequencing approaches have further reinforced this limitation by identifying multiple sub-clones in individual tumour cells (Tang et al., 2021). Second, LBs provide a less invasive, cost effective tool for repeatedly monitoring tumours in comparison to regular tissue biopsies, acquired through medical procedures (Morris and Strickler, 2021). Moreover, experimental accessibility to tissue biopsies is generally limited, creating a diagnostic dilemma, for which LBs can intuitively provide a solution. There is excellent versatility in LBs for both cellular and intra-cellular biomarkers, with bioinformatics and machine learning approaches becoming integral to their rapid advancement in utilization.

Next generation sequencing is a driver for medical liquid biopsy research.

The emergence of LBs is correlated to the increased accessibility of next generation sequencing (NGS) techniques in the medical context. In the next decade, sequencing technologies are predicted to evolve in accordance with the predicted doubling of computing power (Moore’s law) and will continue to influence the viability of LB applications (Wetterstrand, 2022).

One form of NGS is whole genome sequencing (WGS) which provides researchers with expansive information on mutations, copy number variations (CNVs) and insertions and deletions (indels). The high throughput nature of WGS has led to several large scale, pan-cancer databases which are beneficial within medical research, with examples including 'The 100,000 Genomes Project' (Trotman et al., 2022) and 'The Cancer Genome Atlas' (Chang et al., 2013). The two main categories of NGS are short and long read sequencing. Short read sequencing provides accurate interpretation of small genetic fragments and is currently dominated by Ilumina. The length of these fragments vary between different sequencing technologies with the MiSeq size being 2x300bp. Long read sequencing provides a less accurate method for analysing longer fragments of 10KB or more, which performs better in the analysis of repetitive non-coding DNA and epigenetics (Hu et al., 2021). The two main long read sequencing companies are Oxford Nanopore and PacBio.

There is an intricate balance to be found between cost, depth of sequencing and the discovery potential of any cancer sequencing assay (Figure 1). For instance, a small targeted gene panel is more clinically accessible than a high depth WGS or multi-omic approach, but the smaller assay could lack adequate sensitivity for intra-tumour heterogeneity or significant epigenetic arrangements that a high depth untargeted assay can provide (Donoghue et al., 2020). The higher depth of a WGS also requires more advanced computation and analysis, due to more variation and sequencing noise encountered. There are exciting future implications with short and long read technologies complementing each other well to provide genomic and chromosomal mapping (Schöpflin et al., 2022).

The year 2023 is predicted to see a disruption to the sequencing market with added competition being introduced. Ultima genomics have groundbreakingly announced technology for the long sought after $100 genome (Pollie, 2023). Furthermore, PacBio have also announced the new Revio system which will attempt to revolutionise long read sequencing with a $1000 price point (Kovaka et al., 2023). It will be interesting to see how Ilumina's plans of expansion into the long read market integrate with their existing hardware to benefit medical sequencing strategies. In addition, when taking into consideration the advancements in single cell sequencing and spatial-genomics which allow a detailed presentation of gene regulation within tissues, it is conclusive that sequencing technology is not the limitation to further advancement of precision oncology (Rao et al., 2021). Ultimately the increased competition and cost reductions will make large scale clinical sequencing initiatives more financially viable and accessible. However, because NGS allows for flexible analysis of a wide range of biomarkers deciding what to analyse and interpreting this information with high sensitivity and specificity is the significant challenge.

Cell free DNA and other biomarkers show potential for less invasive cancer profiling

Although cell free DNA (cfDNA) is the primary focus of this review, it is important to acknowledge other useful biomarkers. For example, circulating tumour cells (CTCs) have shown associations with late stage metastasis in CRC (Zhao et al., 2017), and have a significantly greater proportion of clinical trials taking place in comparison to cfDNA (NLM, 2023). Additionally, exosomes provide a fascinating insight into the various stages of tumour formation, vascularisation and dissemination, with the potential to rival and complement cfDNA analysis (Cisneros-Villanueva et al., 2022). Exosomes and CTCs have the benefit of containing RNA, DNA and proteins, with multi-analyte investigations alongside cfDNA proving successful (Krug et al., 2018). Exosomes have the added benefit of performing better for early detection compared to ctDNA, which occurs at low quantities during early stage cancer disease. A summary compiled from public clinical trials maintained by the National Library of Medicine, provides a good comparison of the clinical trial frequency for different biomarkers (Table 1).

Recently, attention has also shifted to the cfDNA analyte because it is easier to isolate than CTCs. cfDNA is generated by necrosis and apoptosis by both healthy and pathological cells, its interactions with various parameters such as tissue pathology and cellular senescence affect its presentation within the blood (Rostami et al., 2020).

It has been established that ctDNA can be detected to a high sensitivity and specificity from cfDNA (Page et al., 2017). Large scale sequencing has also confirmed that cfDNA profiles can correlate well with tissue samples and improve clinical outcomes for patients (Zill et al., 2016). Consequently, extensive research is ongoing to support cfDNA as a biomarker for detecting ctDNA in renal cancer (Smith et al., 2020), and CRC (Parikh et al., 2019). Large pan-cancer studies prove that a mutational overview can be identified using ctDNA profiling. However, different cancer types vary in their sensitivity for detection and in a Chinese cohort, differences in blood tumour burden (bTMB) between tissue sequencing and ctDNA, warrant further investigation (Zhang et al., 2021).

Personalised therapeutics by genomic profiling.

One important factor in personalised medicine is individualised patient treatment plans based on their respective biomarker presentation. BC is a model in how treatment response can be guided by cancer phenotype. Tumour markers such as estrogen (ER), progesterone (PR) receptors and human epidermal growth factor receptor 2 (HER2) have been identified, with tumour marker positive and negative patients requiring divergent treatment pathways (Van Poznak et al., 2015). In practice it has been established that genomic analysis is used to support these clinical treatment pathways. It is reasonable to suggest this approach can

be applied to other cancer types, and that detection of ctDNA can contribute to future guidance of personalised treatments.

Another personalised aspect is the monitoring of minimal residual disease (MRD), defined as the small quantities of cancer cells leftover after treatment which are difficult to identify with original diagnostic scans. This is an important healthcare application for ctDNA detection and being able to reliably predict recurrence rates of solid tumours has been demonstrated in CRC (Tie et al., 2016) Furthermore, in lung cancer post treatment ctDNA analysis has been shown to out perform radiology examinations by 72% for MRD detection (20 patients, 255 samples), reducing the median detection time by 5 months (Chaudhuri et al., 2017). In this study, 50% of the cohort had mutation profiles with promising responses to immune checkpoint blockade and tyrosine kinase inhibitors. These biological mechanisms are important checkpoints in cancer progression and are further evidence of the potential efficacy. Another interesting area to consider is the quantification of ctDNA overtime to monitor a patients treatment response (Hsu et al., 2018). Again, this area of research demonstrates that next generation sequencing (NGS) mutation panels can be used to improve therapeutic response and identify genetic variants for metastatic tumours.

cfDNA has an epigenetic fingerprint

While genetic analysis of cfDNA and ctDNA is important, epigenetic factors such as DNA methylation and cfDNA fragmentomics provide another opportunity to interrogate the cancer genome without detecting ctDNA (Lo et al., 2021). Human fragmentation patterns are determined by several factors such as nucleosome positioning, chromatin accessibility, histone modifications and enhancer-promotor interactions. The fragmentation patterns of cfDNA in humans display variation in size, but there are known central tendencies for a range of biological parameters, with healthy individuals having a threshold size of 166bp. This fragmentomic approach was initially identified to distinguish foetal DNA from maternal DNA during pregnancy. It was a proof of concept study in the non-invasive quantification of alleles present for prenatal mutations in the blood disease β-thalassemia. Consequently, it was also possible to generate genomic mapping using the maternal haplotype and paternal genotype (Lo et al., 2010).

Interestingly, cancer cells exhibit smaller fragmentation sizes of 144bp which can be used as a target for the detection of cancer phenotypes. Pan-cancer studies have shown that different types of cancer have a tendency towards different fragment lengths with a 90-150bp limit being distinguished (Mouliere et al., 2018). Evidently fragment length distribution is an important target for a range of disease phenotypes and requires further investigation. Furthermore, chromatin accessibility also varies between cells and correlates with gene regulatory elements and transcription factors within the cell cycle. Histone modifications such

as methylation and acylation show enrichment patterns for the active and silent genes impacting DNA polymerase 2 binding affinity (Carter and Zhao, 2021). These types of analysis can be done via bulk or single cell sequencing methods, with the latter providing a more precise level of epigenetic marker enrichment for heterogenous and rare cell populations (Cusanovich et al., 2015).

DNA methylation is another area of epigenetics gaining traction in a wide range of cancer types, resulting in assays for the detection of hypomethylated sites and CNVs (Chan et al., 2013). Methylation modifications are often found in cancer cells, therefore research into their detection at low sequencing depth would provide efficient use of clinical diagnostic resources. Additionally, genome-wide DNA co-methylation patterns have been shown to display high co-ordination levels between different CpG sites. The CpG location analysis has determined that the tissue of origin can be mapped via cfDNA defined methylated haplotype blocks, which in clinical practice could yield significant implications (Guo et al., 2017). To expand methylation utility, recent proof of concept studies investigating early stage multi-cancer detection have performed well compared to the common genetic mutation panels (Liu et al., 2020). This is advantageous because as mentioned, ctDNA levels in early stages of certain types of cancer result in LBs with low tumour fraction. The innovative use of methylation analysis by Liu et al. (2020) utilises the circulating cell-free genome atlas (CCCA) created by the GRAIL study, which provides an essential resource to study gene regulation using liquid biopsies (Loyfer et al., 2023). The GRAIL study is a large-scale clinical trial of over 300,000 participants from the UK and America that is investigating its early detection technology Galleri. This research collaboration has led to several significant research studies on cancer methylation screening and the disparities caused by factors such as race, sex and socioeconomic status (Tang et al., 2023).

Precision oncology and existing bioinformatics approaches for tumour sub typing

A wide range of bioinformatics approaches are actively being investigated to analyse a diverse range of multi-omic cancer data (summarised in Table 2.0). These computational tools are often used in combination with each other to prepare and analyse cfDNA and ctDNA. This requires specialist expertise to successfully integrate advanced software into reusable bioinformatics pipelines, with pre-processed inputs to obtain the desired outputs (summarised in figure 2). An example output is tumour fraction, which can be defined as the quantity of ctDNA present in a cfDNA sample, and it is agreed that it highly correlates with high tumour burden. IchorCNA is a fairly common tool within the literature used to estimate the proportion of ctDNA within cfDNA samples (Adalsteinsson et al., 2017). Interestingly, it has been used in a diverse range of patient cohorts (Ahuno et al., 2021), but with varying

prediction success as low as 35% in certain studies (Hallermayr et al., 2022). InchorCNA has also been compared by Tsui et al. (2021) to the cf-IMPACT approach with a high positive correlation with genomic wide z-scores generated by shallow WGS and variant allele fractions. Further investigation is required to fully understand why inchorCNA performs differently on different datasets and input parameters.

Transcription factors and their regulation are an important factor in tumourigenesis, and there is a tendency towards bioinformatics pipelines inferring TF accessibility from cfDNA samples. This transcription factor profiling is displayed by Ulz et al. (2019) as a proof of concept in prostate cancer and indicates how non-coding analysis is emerging as a key component of future clinical analysis. Griffin is a similar, more recent approach which demonstrates nucleosomal profiling to predict tumour phenotypes using a novel fragment length GC-bias correction method. It has been designed for ultra low pass WGS x0.1 data, with its best performance being on tumour fractions >= 0.05 (Doebley et al., 2022). Interestingly, it has been able to differentiate transcription factor binding sites (TFBs) and their accessibility in different breast tumour subtypes. Griffin can interpret fragmentation patterns to distinguish nucleosomal positioning. The actively bound TFBs with higher accessibility result in lower sequence coverage when compared to the more protected regions of the genome bound by nucleosomes. Specific regions of interest can be identified beforehand with epigenetic sequencing techniques such as ATAC-Seq and Chip-Seq. In terms of coverage, the prediction AUC scores were lower when coverage was lowered to 0.1x. Other limitations of Griffin include lower performance in tumour fractions below 10% which make it difficult to monitor early detection or MRD. Also, the GC correction method is sensitive to different cfDNA and sequencing approaches limiting its compatibility with other datasets. Further research is required to understand and resolve these limitations. Griffin has potential to become an important tool for tumour phenotype analysis and it has been suggested that it can apply to other transcription regulation mechanisms due to its ability to customise to biologically interesting sites.

Significant variation in sequencing depth, sensitivity and specificity exist

There is a fine balance to be found between higher sequencing depth and its accessibility and viability for routine clinical practice. Within the cancer genomics field there is significant variation in sequencing depth. For example, it has been established that a high sequencing depth provides an ultra-sensitive method for ctDNA detection from cfDNA by using the multi-phase bioinformatics tool called CAPP-Seq (Newman et al., 2014). This approach uses population analysis of recurrent mutations to generate a CAPP-Seq library which can then be used in patient level analysis. There is variation in the sensitivity (true positive) and

specificity (true negative) between inchorCNA, Griffin and CAPP-Seq. When you take into consideration pan-cancer variability in tumour fraction, significant challenges arise to create a standardised framework for personalised medicine. The current evidence suggests that while these bioinformatics tools are theoretically applicable to multiple cancer types, a universal bioinformatics pipeline or software for precision oncology does not currently exist. This review proposes that it is much more likely that a divergent approach will continue between different cancer types. Biomarkers such as cfDNA are much smaller in size than the whole human genome and research will continue on low depth sequencing samples to determine the maximum clinical insight that can be achieved from minimal datasets.

Advanced multi-omic computational models and machine learning

While evidence has shown significant variability in the bioinformatic pipelines available, there is one factor emerging consistently, the use of advanced computational and machine learning tools to analyse and predict tumour phenotypes. Cancer genomics is in a transitional period where specialised algorithms that process large amounts of data are used to contextualise multi-omic information. This has resulted in several challenges for researchers and a demand for specialist genomic computing infrastructure. Multi-omics is still in its infancy but interesting multimodal approaches are emerging with highly accurate results. With the correct data structures in place it is possible to analyse genomic, pathology and radiology data to predict treatment responses in patients (Vanguri et al., 2022). Similarly, mathematical models have been created which provide insight on the sub-clonal evolution of cancers. These aim to use genomic data and population genetics to provide an insightful quantitative framework to predict the trajectories of tumours (Williams et al., 2018). Most studies claim high performance on their experimental datasets, but caution is required to ensure results are reproducible. While it has been presented that multi-analyte analysis can lead to better inference than individual biomarkers, it is difficult to fully standardise the different healthcare inputs required. This is one reason why cfDNA approaches are being investigated for a wide range of applications.

Liquid biopsies and cell free DNA provide limited complementary diagnostics

Liquid biopsies, cfDNA and ctDNA detection currently provide a complementary tool to traditional cancer diagnostics tests and imaging. They have considerable potential, and the wide range of published evidence highlights the applications of how cfDNA is already impacting clinical care. However, it would be premature to assume that the precision oncology revolution has been achieved, because the aim of replacing invasive tissue biopsies requires key milestones to be overcome. A good example of this is how most pre-clinical studies are still comparing results against tissue biopsies standards with varying

degrees of success (Zhang et al., 2021). There are several limitations and bottlenecks that need to be overcome to implement personalised oncology globally. Firstly, the literature does not present a clear bioinformatics framework to achieve the best sensitivity and specificity for each cancer type. There is a lack of standardisation in factors such as the minimal coverage threshold required and which bioinformatics approach is best suited. Mutation panels are more clinically applicable than more advanced WGS methods investigating epigenetic and structural tumour phenotypes, with approaches like PanelApp inferring variant priority for panels virtually from WGS data (Martin et al., 2019). Methods to identify epigenetic fingerprints cannot be ignored and require the same clinical utility as the more accessible gene panels, because tumour heterogeneity is majorly determined by epigenetic regulation. Large collaborative projects such as the previously mentioned GRAIL study are actively researching the non-coding areas of the genome (Tang et al., 2023). Numerous studies presented here use relatively small sample sizes compared to cancer cases, and these bioinformatics approaches need to be applied on a population-wide scale with normal controls to ensure maintenance of specificity and sensitivity. By its nature, cancer is a disease of variation, and it is unlikely with the current technology that a breakout analysis pipeline will be applicable to all cancer phenotypes or sequencing approaches.

Secondly, there is a significant issue of collection and diversification of existing genomic databases and experimental data. It is acknowledged within the genomic literature that there is European bias to genome-wide association studies (Conti et al., 2021), with a wide range of new genetic loci being identified when focusing on African ancestry in CRC (Wang et al., 2014). Additionally, an investigation into a promising BC screening method has shown significant over estimation in non-European groups, indicating a separate test is required for non-European patients (Evans et al., 2022). The complexity of inferred ancestral origin is a significant challenge for mutant allele detection and requires improvement within precision medicine (Pereira et al., 2021). Large amounts of variation occur at a sub-continental level in Africa, for instance East and West African participants display significant differences in analysis output from existing bioinformatics pipelines. It is essential that these limitations are overcome to provide an equitable personalised medicine system. Only then can clinicians have full confidence in genetic tests that perform to the same standard for patients with different ancestral origins. Also, there needs to be an up-skilling of existing healthcare staff to analyse and interpret the vast amounts of genomic data being collected. Sequencing technology is consistently improving creating more precise and expansive landscapes of the genome, but there is a shortage of skilled professionals to understand this data. Computationally, it is common for organisations to not have sufficient structures in place to incorporate large datasets into their research.

cfDNA does require quite a labour-intensive protocol to isolate and this does occur in other biomarkers, such as CTCs which requires improvement and automation. There is real competition between the different biomarkers within precision oncology, and it is arguable that all the different types covered in this review can be effective in different disease presentations. Early detection of cancer is another limitation of ctDNA specifically and an area that can be improved on by epigenetic studies discussed such as Liu et al. (2020). Arguably, an aim for genomic medicine is to achieve earlier detection of cancer, and reduce the prevalence of advanced metastatic cancer types.

In conclusion, LBs and cfDNA analysis will be a high priority for cancer diagnostics and therapeutics research for the foreseeable future. There are key challenges that need to be overcome and it is likely that a range of different biomarkers will be used to improve therapeutic outcomes for a diverse range of patients. Experimental analysis of bioinformatic pipelines such as Griffin are required on a wide range of different cancer phenotypes and patient cohorts. The software and study supporting Griffin suggest that it is widely applicable to other cancer types and gene regulation mechanisms (Doebley et al., 2022). However, there is limited experimental evidence on how Griffin's specificity and sensitivity perform in CRC patients, as well as other cancer types. Parameters such as tumour fraction, cancer stage, sequencing coverage and patient ancestry all require an in depth assessment. Further research and software development is required to understand and improve Griffins workflow compatibility issues, which will lead to an even more versatile tool for GC bias and nucleosome profiling. It is feasible that pipeline results will vary between cancer types and further enhancement of the software will be required in the future.
Additionally, these parameters need to be thoroughly compared to similar experimental bioinformatic pipelines in development, reporting high performance for other biomarkers such as histone modifications and cell free nucleosomes (Fedyuk et al., 2023). Through experimental investigation, it will be possible for software like Griffin to be incorporated into future clinical bioinformatics pipelines. This work is important to close the gap in clinical feasibility that exists between epigenetic marker assays and the more common gene mutation panels. Society is rapidly moving towards a digital healthcare system which collects genetic and personal information to provide inferences on future health outcomes. LBs and cfDNA provide a non-invasive snapshot of a patients health profile, made accessible by the computational processing of a patients most personal assets, the genome and epigenome.

References

Adalsteinsson, V. A., Ha, G., Freeman, S. S., Choudhury, A. D., Stover, D. G., Parsons, H. A., Gydush, G., Reed, S. C., Rotem, D., Rhoades, J., Loginov, D., Livitz, D., Rosebrock, D., Leshchiner, I., Kim, J., Stewart, C., Rosenberg, M., Francis, J. M., Zhang, C.-Z., Cohen, O., Oh, C., Ding, H., Polak, P., Lloyd, M., Mahmud, S., Helvie, K., Merrill, M. S., Santiago, R. A., O'connor, E. P., Jeong, S. H., Leeson, R., Barry, R. M., Kramkowski, J. F., Zhang, Z., Polacek, L., Lohr, J. G., Schleicher, M., Lipscomb, E., Saltzman, A., Oliver, N. M., Marini, L., Waks, A. G., Harshman, L. C., Tolaney, S. M., Van Allen, E. M., Winer, E. P., Lin, N. U., Nakabayashi, M., Taplin, M.-E., Johannessen, C. M., Garraway, L. A., Golub, T. R., Boehm, J. S., Wagle, N., Getz, G., Love, J. C. & Meyerson, M. 2017. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nature Communications,* 8**,** 1324.

Ahuno, S. T., Doebley, A.-L., Ahearn, T. U., Yarney, J., Titiloye, N., Hamel, N., Adjei, E., Clegg-Lamptey, J.-N., Edusei, L. & Awuah, B. 2021. Circulating tumor DNA is readily detectable among ghanaian breast cancer patients supporting non-invasive cancer genomic studies in africa. *NPJ precision oncology,* 5**,** 1-8.

Carter, B. & Zhao, K. 2021. The epigenetic basis of cellular heterogeneity. *Nat Rev Genet,* 22**,** 235-250.

Chan, K. A., Jiang, P., Chan, C. W., Sun, K., Wong, J., Hui, E. P., Chan, S. L., Chan, W. C., Hui, D. S. & Ng, S. S. 2013. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proceedings of the National Academy of Sciences,* 110**,** 18761-18768.

Chang, K., Creighton, C. J., Davis, C., Donehower, L., Drummond, J., Wheeler, D., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y. S. N., Chu, A., Chuah, E., Chun, H.-J. E., Dhalla, N., Guin, R., Hirst, M., Hirst, C., Holt, R. A., Jones, S. J. M., Lee, D., Li, H. I., Marra, M. A., Mayo, M., Moore, R. A., Mungall, A. J., Robertson, A. G., Schein, J. E., Sipahimalani, P., Tam, A., Thiessen, N., Varhol, R. J., Beroukhim, R., Bhatt, A. S., Brooks, A. N., Cherniack, A. D., Freeman, S. S., Gabriel, S. B., Helman, E., Jung, J., Meyerson, M., Ojesina, A. I., Pedamallu, C. S., Saksena, G., Schumacher, S. E., Tabak, B., Zack, T., Lander, E. S., Bristow, C. A., Hadjipanayis, A., Haseley, P., Kucherlapati, R., Lee, S., Lee, E., Luquette, L. J., Mahadeshwar, H. S., Pantazi, A., Parfenov, M., Park, P. J., Protopopov, A., Ren, X., Santoso, N., Seidman, J., Seth, S., Song, X., Tang, J., Xi, R., Xu, A. W., Yang, L., Zeng, D., Auman, J. T., Balu, S., Buda, E., Fan, C., Hoadley, K. A., Jones, C. D., Meng, S., Mieczkowski, P. A., Parker, J. S., Perou, C. M., Roach, J., Shi, Y., Silva, G. O., Tan, D., Veluvolu, U.,

Waring, S., Wilkerson, M. D., Wu, J., Zhao, W., Bodenheimer, T., Hayes, D. N., Hoyle, A. P., Jeffreys, S. R., Mose, L. E., Simons, J. V., Soloway, M. G., Baylin, S. B., Berman, B. P., Bootwalla, M. S., Danilova, L., Herman, J. G., et al. 2013. The cancer genome atlas pan-cancer analysis project. *Nature Genetics,* 45**,** 1113-1120.

Chaudhuri, A. A., Chabon, J. J., Lovejoy, A. F., Newman, A. M., Stehr, H., Azad, T. D., Khodadoust, M. S., Esfahani, M. S., Liu, C. L. & Zhou, L. 2017. Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer discovery,* 7**,** 1394-1403.

Cisneros-Villanueva, M., Hidalgo-Pérez, L., Rios-Romero, M., Cedro-Tanda, A., Ruiz-Villavicencio, C., Page, K., Hastings, R., Fernandez-Garcia, D., Allsopp, R. & Fonseca-Montaño, M. 2022. Cell-free DNA analysis in current cancer clinical trials: A review. *British Journal of Cancer,* 126**,** 391-400.

Conti, D. V., Darst, B. F., Moss, L. C., Saunders, E. J., Sheng, X., Chou, A., Schumacher, F. R., Olama, A. a. A., Benlloch, S., Dadaev, T., Brook, M. N., Sahimi, A., Hoffmann, T. J., Takahashi, A., Matsuda, K., Momozawa, Y., Fujita, M., Muir, K., Lophatananon, A., Wan, P., Le Marchand, L., Wilkens, L. R., Stevens, V. L., Gapstur, S. M., Carter, B. D., Schleutker, J., Tammela, T. L. J., Sipeky, C., Auvinen, A., Giles, G. G., Southey, M. C., Macinnis, R. J., Cybulski, C., Wokołorczyk, D., Lubiński, J., Neal, D. E., Donovan, J. L., Hamdy, F. C., Martin, R. M., Nordestgaard, B. G., Nielsen, S. F., Weischer, M., Bojesen, S. E., Røder, M. A., Iversen, P., Batra, J., Chambers, S., Moya, L., Horvath, L., Clements, J. A., Tilley, W., Risbridger, G. P., Gronberg, H., Aly, M., Szulkin, R., Eklund, M., Nordström, T., Pashayan, N., Dunning, A. M., Ghoussaini, M., Travis, R. C., Key, T. J., Riboli, E., Park, J. Y., Sellers, T. A., Lin, H.-Y., Albanes, D., Weinstein, S. J., Mucci, L. A., Giovannucci, E., Lindstrom, S., Kraft, P., Hunter, D. J., Penney, K. L., Turman, C., Tangen, C. M., Goodman, P. J., Thompson, I. M., Hamilton, R. J., Fleshner, N. E., Finelli, A., Parent, M.-É., Stanford, J. L., Ostrander, E. A., Geybels, M. S., Koutros, S., Freeman, L. E. B., Stampfer, M., Wolk, A., Håkansson, N., Andriole, G. L., Hoover, R. N., Machiela, M. J., Sørensen, K. D., Borre, M., Blot, W. J., Zheng, W., Yeboah, E. D., Mensah, J. E., Lu, Y.-J., et al. 2021. Trans-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. *Nature Genetics,* 53**,** 65-75.

Coombes, R. C., Page, K., Salari, R., Hastings, R. K., Armstrong, A., Ahmed, S., Ali, S., Cleator, S., Kenny, L. & Stebbing, J. 2019. Personalized detection of circulating tumor DNA antedates breast cancer metastatic recurrencepersonalized ctDNA detection of breast cancer recurrence. *Clinical Cancer Research,* 25**,** 4255-4263.

Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C. & Shendure, J. 2015. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science,* 348**,** 910-914.

Doebley, A.-L., Ko, M., Liao, H., Cruikshank, A. E., Santos, K., Kikawa, C., Hiatt, J. B., Patton, R. D., De Sarkar, N. & Collier, K. A. 2022. A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. *Nature Communications,* 13**,** 1-18.

Donoghue, M. T., Schram, A. M., Hyman, D. M. & Taylor, B. S. 2020. Discovery through clinical sequencing in oncology. *Nature cancer,* 1**,** 774-783.

Evans, D. G., Van Veen, E. M., Byers, H., Roberts, E., Howell, A., Howell, S. J., Harkness, E. F., Brentnall, A., Cuzick, J. & Newman, W. G. 2022. The importance of ethnicity: Are breast cancer polygenic risk scores ready for women who are not of white european origin? *International journal of cancer,* 150**,** 73-79.

Fedyuk, V., Erez, N., Furth, N., Beresh, O., Andreishcheva, E., Shinde, A., Jones, D., Zakai, B. B., Mavor, Y., Peretz, T., Hubert, A., Cohen, J. E., Salah, A., Temper, M., Grinshpun, A., Maoz, M., Zick, A., Ron, G. & Shema, E. 2023. Multiplexed, single-molecule, epigenetic analysis of plasma-isolated nucleosomes for cancer diagnostics. *Nature Biotechnology,* 41**,** 212-221.

Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A. & Tarpey, P. 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl j Med,* 366**,** 883-892.

Guo, S., Diep, D., Plongthongkum, N., Fung, H.-L., Zhang, K. & Zhang, K. 2017. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nature genetics,* 49**,** 635-642.

Hallermayr, A., Wohlfrom, T., Steinke-Lange, V., Benet-Pagès, A., Scharf, F., Heitzer, E., Mansmann, U., Haberl, C., De Wit, M. & Vogelsang, H. 2022. Somatic copy number alteration and fragmentation analysis in circulating tumor DNA for cancer screening and treatment monitoring in colorectal cancer patients. *Journal of Hematology & Oncology,* 15**,** 1-14.

Heidrich, I., Ačkar, L., Mossahebi Mohammadi, P. & Pantel, K. 2021. Liquid biopsies: Potential and challenges. *International journal of cancer,* 148**,** 528-545.

Hsu, H.-C., Lapke, N., Wang, C.-W., Lin, P.-Y., You, J. F., Yeh, C. Y., Tsai, W.-S., Hung, H. Y., Chiang, S.-F., Chen, H.-C., Chen, S.-J., Hsu, A. & Yang, T. S. 2018. Targeted sequencing of circulating tumor DNA to monitor genetic variants and therapeutic response in metastatic colorectal cancer. *Molecular Cancer Therapeutics,* 17**,** 2238-2247.

Hu, T., Chitnis, N., Monos, D. & Dinh, A. 2021. Next-generation sequencing technologies: An overview. *Human Immunology,* 82**,** 801-811.

Ignatiadis, M., Sledge, G. W. & Jeffrey, S. S. 2021. Liquid biopsy enters the clinic—implementation issues and future challenges. *Nature reviews Clinical oncology,* 18**,** 297-312.

Kovaka, S., Ou, S., Jenike, K. M. & Schatz, M. C. 2023. Approaching complete genomes, transcriptomes and epi-omes with accurate long-read sequencing. *Nature Methods,* 20**,** 12-16.

Krug, A., Enderle, D., Karlovich, C., Priewasser, T., Bentink, S., Spiel, A., Brinkmann, K., Emenegger, J., Grimm, D. & Castellanos-Rizaldos, E. 2018. Improved EGFR mutation detection using combined exosomal RNA and circulating tumor DNA in NSCLC patient plasma. *Annals of Oncology,* 29**,** 700-706.

Liu, M., Oxnard, G., Klein, E., Swanton, C., Seiden, M., Liu, M. C., Oxnard, G. R., Klein, E. A., Smith, D. & Richards, D. 2020. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology,* 31**,** 745-759.

Lo, Y. D., Chan, K. A., Sun, H., Chen, E. Z., Jiang, P., Lun, F. M., Zheng, Y. W., Leung, T. Y., Lau, T. K. & Cantor, C. R. 2010. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Science translational medicine,* 2**,** 61ra91-61ra91.

Lo, Y. D., Han, D. S., Jiang, P. & Chiu, R. W. 2021. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science,* 372**,** eaaw3616.

Loyfer, N., Magenheim, J., Peretz, A., Cann, G., Bredno, J., Klochendler, A., Fox-Fisher, I., Shabi-Porat, S., Hecht, M., Pelet, T., Moss, J., Drawshy, Z., Amini, H., Moradi, P., Nagaraju, S., Bauman, D., Shveiky, D., Porat, S., Dior, U., Rivkin, G., Or, O., Hirshoren, N., Carmon, E., Pikarsky, A., Khalaileh, A., Zamir, G., Grinbaum, R., Abu Gazala, M., Mizrahi, I., Shussman, N., Korach, A., Wald, O., Izhar, U., Erez, E., Yutkin, V., Samet, Y., Rotnemer Golinkin, D., Spalding, K. L., Druid, H., Arner, P., Shapiro, A. M. J., Grompe, M., Aravanis, A., Venn, O., Jamshidi, A., Shemer, R., Dor, Y., Glaser, B. & Kaplan, T. 2023. A DNA methylation atlas of normal human cell types. *Nature,* 613**,** 355-364.

Martin, A. R., Williams, E., Foulger, R. E., Leigh, S., Daugherty, L. C., Niblock, O., Leong, I. U. S., Smith, K. R., Gerasimenko, O., Haraldsdottir, E., Thomas, E., Scott, R. H., Baple, E., Tucci, A., Brittain, H., De Burca, A., Ibañez, K., Kasperaviciute, D., Smedley, D., Caulfield, M., Rendon, A. & Mcdonagh, E. M. 2019. Panelapp crowdsources expert knowledge to establish consensus diagnostic gene panels. *Nature Genetics,* 51**,** 1560-1565.

555 Morris, V. K. & Strickler, J. H. 2021. Use of circulating cell-free DNA to guide precision
556 medicine in patients with colorectal cancer. *Annual Review of Medicine,* 72**,** 399-413.
557 Mouliere, F., Chandrananda, D., Piskorz, A. M., Moore, E. K., Morris, J., Ahlborn, L. B., Mair,
558 R., Goranova, T., Marass, F. & Heider, K. 2018. Enhanced detection of circulating
559 tumor DNA by fragment size analysis. *Science translational medicine,* 10**,** eaat4921.
560 Newman, A. M., Bratman, S. V., To, J., Wynne, J. F., Eclov, N. C., Modlin, L. A., Liu, C. L.,
561 Neal, J. W., Wakelee, H. A. & Merritt, R. E. 2014. An ultrasensitive method for
562 quantitating circulating tumor DNA with broad patient coverage. *Nature medicine,* 20**,**
563 548-554.
564 Nlm. 2023. *Clinicaltrials.Gov* [Online]. Available: https://clinicaltrials.gov/ [Accessed
565 19/02/2023].
566 Otandault, A., Anker, P., Dache, Z. a. A., Guillaumon, V., Meddeb, R., Pastor, B., Pisareva,
567 E., Sanchez, C., Tanos, R. & Tousch, G. 2019. Recent advances in circulating
568 nucleic acids in oncology. *Annals of Oncology,* 30**,** 374-384.
569 Page, K., Guttery, D. S., Fernandez-Garcia, D., Hills, A., Hastings, R. K., Luo, J., Goddard,
570 K., Shahin, V., Woodley-Barker, L., Rosales, B. M., Coombes, R. C., Stebbing, J. &
571 Shaw, J. A. 2017. Next generation sequencing of circulating cell-free DNA for
572 evaluating mutations and gene amplification in metastatic breast cancer. *Clinical*
573 *Chemistry,* 63**,** 532-541.
574 Parikh, A. R., Leshchiner, I., Elagina, L., Goyal, L., Levovitz, C., Siravegna, G., Livitz, D.,
575 Rhrissorrakrai, K., Martin, E. E. & Van Seventer, E. E. 2019. Liquid versus tissue
576 biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal
577 cancers. *Nature medicine,* 25**,** 1415-1421.
578 Pereira, L., Mutesa, L., Tindana, P. & Ramsay, M. 2021. African genetic diversity and
579 adaptation inform a precision medicine agenda. *Nature Reviews Genetics,* 22**,** 284-
580 306.
581 Pollie, R. 2023. Genomic sequencing costs set to head down again. Elsevier.
582 Rao, A., Barkley, D., França, G. S. & Yanai, I. 2021. Exploring tissue architecture using
583 spatial transcriptomics. *Nature,* 596**,** 211-220.
584 Rostami, A., Lambie, M., Yu, C. W., Stambolic, V., Waldron, J. N. & Bratman, S. V. 2020.
585 Senescence, necrosis, and apoptosis govern circulating cell-free DNA release
586 kinetics. *Cell Rep,* 31**,** 107830.
587 Schöpflin, R., Melo, U. S., Moeinzadeh, H., Heller, D., Laupert, V., Hertzberg, J., Holtgrewe,
588 M., Alavi, N., Klever, M.-K. & Jungnitsch, J. 2022. Integration of Hi-C with short and
589 long-read genome sequencing reveals the structure of germline rearranged
590 genomes. *Nature Communications,* 13**,** 6470.

Smith, C. G., Moser, T., Mouliere, F., Field-Rayner, J., Eldridge, M., Riediger, A. L., Chandrananda, D., Heider, K., Wan, J. C. M., Warren, A. Y., Morris, J., Hudecova, I., Cooper, W. N., Mitchell, T. J., Gale, D., Ruiz-Valdepenas, A., Klatte, T., Ursprung, S., Sala, E., Riddick, A. C. P., Aho, T. F., Armitage, J. N., Perakis, S., Pichler, M., Seles, M., Wcislo, G., Welsh, S. J., Matakidou, A., Eisen, T., Massie, C. E., Rosenfeld, N., Heitzer, E. & Stewart, G. D. 2020. Comprehensive characterization of cell-free tumor DNA in plasma and urine of patients with renal tumors. *Genome Medicine,* 12.

Tang, J., Tu, K., Lu, K., Zhang, J., Luo, K., Jin, H., Wang, L., Yang, L., Xiao, W., Zhang, Q., Liu, X., Ge, X. Y., Li, G., Zhou, Z. & Xie, D. 2021. Single-cell exome sequencing reveals multiple subclones in metastatic colorectal carcinoma. *Genome Med,* 13**,** 148.

Tang, W. W., Yimer, H., Tummala, M., Shao, S., Chung, G., Clement, J., Chu, B. C., Hubbell, E., Kurtzman, K. N. & Swanton, C. 2023. Performance of a targeted methylation-based multi-cancer early detection test by race and ethnicity. *Preventive Medicine,* 167**,** 107384.

Tie, J., Wang, Y., Tomasetti, C., Li, L., Springer, S., Kinde, I., Silliman, N., Tacey, M., Wong, H.-L. & Christie, M. 2016. Circulating tumor DNA analysis detects minimal residual disease and predicts recurrence in patients with stage ii colon cancer. *Science translational medicine,* 8**,** 346ra92-346ra92.

Trotman, J., Armstrong, R., Firth, H., Trayers, C., Watkins, J., Allinson, K., Jacques, T. S., Nicholson, J. C., Burke, G. a. A., Behjati, S., Murray, M. J., Hook, C. E., Tarpey, P. & Genomics England Res, C. 2022. The NHS England 100,000 genomes project: Feasibility and utility of centralised genome sequencing for children with cancer. *British Journal of Cancer,* 127**,** 137-144.

Tsui, D. W. Y., Cheng, M. L., Shady, M., Yang, J. L., Stephens, D., Won, H., Srinivasan, P., Huberman, K., Meng, F., Jing, X., Patel, J., Hasan, M., Johnson, I., Gedvilaite, E., Houck-Loomis, B., Socci, N. D., Selcuklu, S. D., Seshan, V. E., Zhang, H., Chakravarty, D., Zehir, A., Benayed, R., Arcila, M., Ladanyi, M., Funt, S. A., Feldman, D. R., Li, B. T., Razavi, P., Rosenberg, J., Bajorin, D., Iyer, G., Abida, W., Scher, H. I., Rathkopf, D., Viale, A., Berger, M. F. & Solit, D. B. 2021. Tumor fraction-guided cell-free DNA profiling in metastatic solid tumor patients. *Genome Medicine,* 13**,** 96.

Ulz, P., Perakis, S., Zhou, Q., Moser, T., Belic, J., Lazzeri, I., Wölfler, A., Zebisch, A., Gerger, A. & Pristauz, G. 2019. Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. *Nature communications,* 10**,** 4666.

627 Van Der Toom, E. E., Axelrod, H. D., De La Rosette, J. J., De Reijke, T. M., Pienta, K. J. &
628 Valkenburg, K. C. 2019. Prostate-specific markers to identify rare prostate cancer
629 cells in liquid biopsies. *Nature Reviews Urology,* 16**,** 7-22.
630 Van Poznak, C., Somerfield, M. R., Bast, R. C., Cristofanilli, M., Goetz, M. P., Gonzalez-
631 Angulo, A. M., Hicks, D. G., Hill, E. G., Liu, M. C., Lucas, W., Mayer, I. A., Mennel, R.
632 G., Symmans, W. F., Hayes, D. F. & Harris, L. N. 2015. Use of biomarkers to guide
633 decisions on systemic therapy for women with metastatic breast cancer: American
634 society of clinical oncology clinical practice guideline. *J Clin Oncol,* 33**,** 2695-704.
635 Vanguri, R. S., Luo, J., Aukerman, A. T., Egger, J. V., Fong, C. J., Horvat, N., Pagano, A.,
636 Araujo-Filho, J. D. a. B., Geneslaw, L., Rizvi, H., Sosa, R., Boehm, K. M., Yang, S.-
637 R., Bodd, F. M., Ventura, K., Hollmann, T. J., Ginsberg, M. S., Gao, J., Vanguri, R.,
638 Hellmann, M. D., Sauter, J. L., Shah, S. P. & Consortium, M. M. 2022. Multimodal
639 integration of radiology, pathology and genomics for prediction of response to PD-
640 (L)1 blockade in patients with non-small cell lung cancer. *Nature Cancer,* 3**,** 1151-
641 1164.
642 Wang, H., Burnett, T., Kono, S., Haiman, C. A., Iwasaki, M., Wilkens, L. R., Loo, L. W., Van
643 Den Berg, D., Kolonel, L. N. & Henderson, B. E. 2014. Trans-ethnic genome-wide
644 association study of colorectal cancer identifies a new susceptibility locus in VTI1A.
645 *Nature communications,* 5**,** 4613.
646 Wetterstrand, K. 2022. *DNA sequencing costs: Data from the NHGRI genome sequencing*
647 *program (GSP)* [Online]. Available: Available at:
648 www.genome.gov/sequencingcostsdata. [Accessed 19/02/2023 2023].
649 Williams, M. J., Werner, B., Heide, T., Curtis, C., Barnes, C. P., Sottoriva, A. & Graham, T.
650 A. 2018. Quantification of subclonal selection in cancer from bulk sequencing data.
651 *Nature Genetics,* 50**,** 895-903.
652 Zeinali, M., Lee, M., Nadhan, A., Mathur, A., Hedman, C., Lin, E., Harouaka, R., Wicha, M.
653 S., Zhao, L. & Palanisamy, N. 2020. High-throughput label-free isolation of
654 heterogeneous circulating tumor cells and CTC clusters from non-small-cell lung
655 cancer patients. *Cancers,* 12**,** 127.
656 Zhang, Y., Yao, Y., Xu, Y., Li, L., Gong, Y., Zhang, K., Zhang, M., Guan, Y., Chang, L., Xia,
657 X., Li, L., Jia, S. & Zeng, Q. 2021. Pan-cancer circulating tumor DNA detection in
658 over 10,000 chinese patients. *Nature Communications,* 12**,** 11.
659 Zhao, R., Cai, Z., Li, S., Cheng, Y., Gao, H., Liu, F., Wu, S., Liu, S., Dong, Y. & Zheng, L.
660 2017. Expression and clinical relevance of epithelial and mesenchymal markers in
661 circulating tumor cells from colorectal cancer. *Oncotarget,* 8**,** 9293.
662 Zill, O. A., Mortimer, S., Banks, K. C., Nagy, R. J., Chudova, D., Jackson, C., Baca, A., Ye,
663 J. Z., Lanman, R. B., Talasaz, A., Eltoukhy, H. & Kurzrock, R. 2016. Somatic

genomic landscape of over 15,000 patients with advanced-stage cancer from clinical next-generation sequencing analysis of circulating tumor DNA. *Journal of Clinical Oncology,* 34**,** LBA11501-LBA11501.

Figures and Tables

Search Term	Entire Database *
liquid biopsy	439 studies
exosome	299 studies
cfDNA	655 studies
CTCs	3,061 studies

Table 1 – Search results from clinicaltrials.gov (NLM, 2023)

Searching on the clinical trails.gov shows the amount of clinical trials recorded in this database for all status types (e.g active & completed).

* This includes synonyms such as cfDNA/ cell free DNA.

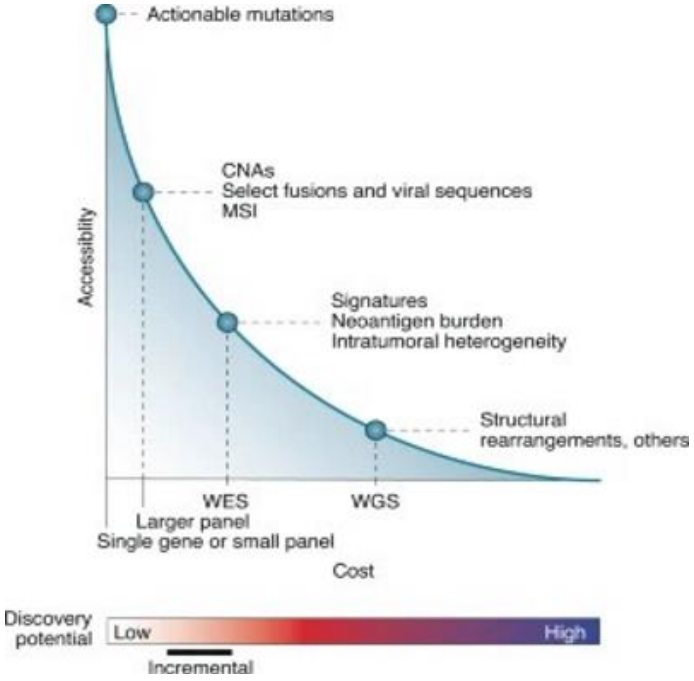

Figure 1 – The sequencing factors influencing discovery potential (Donoghue et al., 2020 ,Fig.1)
Small mutation panels are the most accessible type of NGS cancer assay with a lower cost.
As you start to move into WGS approaches the discovery potential for epigenetic and structural arrangements increases but the cost and accessibility increases.

Bioinformatic approach	Coverage	Genetic / epigenetic	Tumour fraction	Medical potential	Medical Limitations
IchorCNA (Adalsteinsson et al., 2017)	0.1x	Genetic WES, WGS Estimates Tumour fraction	>= 0.03	Sensitivity 0.95 Specificity 0.91 Tumour fraction calculated from low coverage is a viable diagnostic Used on different cfDNA samples	No early detection
Griffin (Doebley et al., 2022)	0.1x	Epigenetic, ChIP-Seq ATAC-seq TFBs and nucleosomal profiling.	>=0.05	Metastatic sub typing Prediction and detection of cancer sub types	No early detection Requires specific standardised cfDNA protocols
TranscriptionFactorProfiling (Ulz et al., 2019)	0.2x	Epigenetic ChIP-Seq ATAC-seq TFBs and nucleosomal profiling.	<0.03	Prediction of Prostate cancer sub types. Potential for early detection in CRC	No subtyping of cancer TFs.
CAPP-Seq (Newman et al., 2014)	10,000x	Genetic, targeted approach to find mutations.	>=0.4	High sensitivity (100%) in ctDNA detection stage II-IV	Low sensitivity (50%) to stage I tumours Cannot detect CNAs

Table 2 – An overview of the different bioinformatics approaches

Emerging technologies such as Griffin and inchorCNA focus on lower depth compared to successful CAPP-Seq approaches.

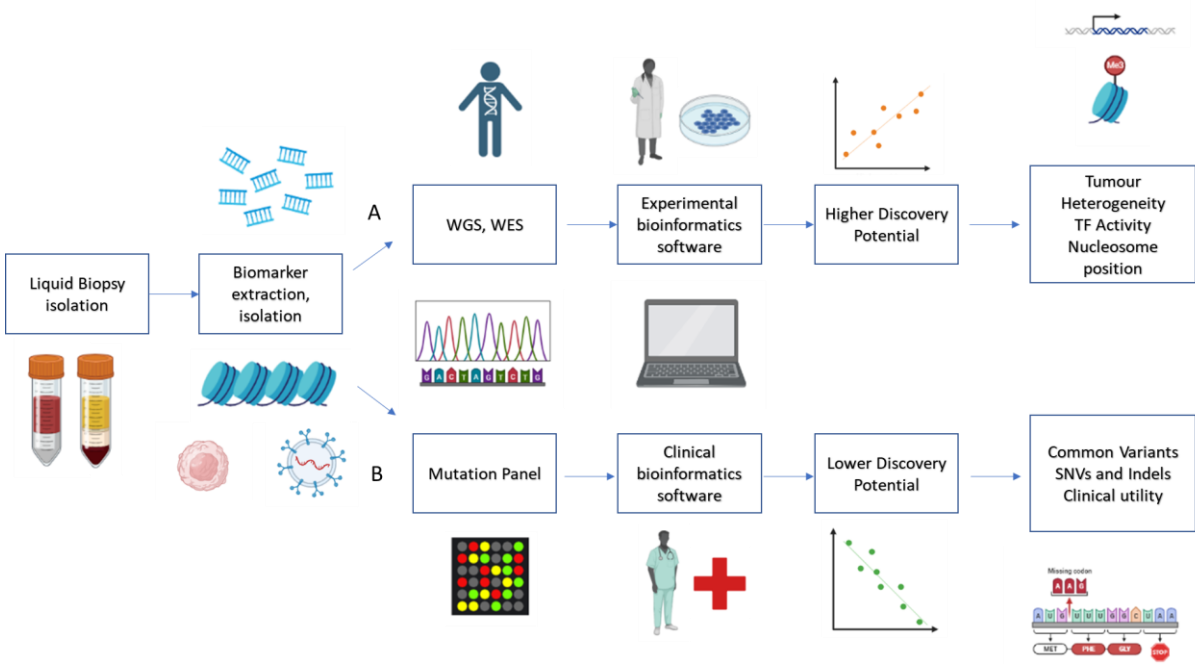

Figure 2 - A simplified liquid biopsy to discovery workflow
Significant work is required at each stage of this process, blood is centrifuged and cfDNA extracted in the required quantity. Sequenced DNA fragments require computational pre-processing before it can be inputted into a clinical or experimental pipeline. Generally mutation panels (B) provide clinical utility while epigenetic analysis (A) is under scientific development with clinical potential. Created with BioRender.com

731

732