Minireview

At the dawn of the transcriptomic medicine

Gea Koks[1], Abigail L Pfaff[2,3], Vivien J Bubb[4], John P Quinn[4] and Sulev Koks[2,3]

[1]Prion Ltd, Tartu 50410, Estonia; [2]Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Murdoch 6150, Australia; [3]Perron Institute for Neurological and Translational Science, Nedlands 6009, Australia; [4]Department of Pharmacology and Therapeutics, University of Liverpool, Liverpool L69 3BX, UK

Corresponding author: Sulev Koks. Email: sulev.koks@perron.uwa.edu.au

Impact statement

This review describes the impact of transcriptomics on experimental biology and its integration into medical practice. Transcriptomics is an essential part of modern biomedical research based on highly sophisticated and reliable technology. Transcriptomics can aid clinical practice and improve the precision of clinical diagnoses and decision-making by complementing existing clinical best practice. The power of which will be increased when combined with genomic variation from genome wide association studies and next generation sequencing. We are witnessing the implementation of RNA-based technologies in clinical practice that will eventually lead to the establishment of transcriptional medicine as a routine tool in diagnosis.

Abstract

Progress in genomic analytical technologies has improved our possibilities to obtain information regarding DNA, RNA, and their dynamic changes that occur over time or in response to specific challenges. This information describes the blueprint for cells, tissues, and organisms and has fundamental importance for all living organisms. This review focuses on the technological challenges to analyze the transcriptome and what is the impact of transcriptomics on precision medicine. The transcriptome is a term that covers all RNA present in cells and a substantial part of it will never be translated into protein but is nevertheless functional in determining cell phenotype. Recent developments in transcriptomics have challenged the fundamentals of the central dogma of biology by providing evidence of pervasive transcription of the genome. Such massive transcriptional activity is challenging the definition of a gene and especially the term "pseudogene" that has now been demonstrated in many examples to be both transcribed and translated. We also review the common sources of biomaterials for transcriptomics and justify the suitability of whole blood RNA as the current optimal analyte for clinical transcriptomics. At the end of the review, a brief overview of the clinical implications of transcriptomics in clinical trial design and clinical diagnosis is given. Finally, we introduce the transcriptome as a target for modern drug development as a tool for extending our capacity for precision medicine in multiple diseases.

Keywords: Transcriptome, RNA-Seq, gene expression profiling, precision medicine, genomics, molecular targeted therapy

***Experimental Biology and Medicine* 2021; 246: 286–292. DOI: 10.1177/1535370220954788**

Introduction

Since the identification of the structure of proteins and nucleic acids and the mechanisms of gene expression, the central concept of biology has underpinned our understanding of gene function.[1] According to this concept, the information in the cell is from DNA to RNA and subsequently translated into proteins. Therefore, the function of genes should be analyzed only by their ability to produce proteins and that proteins define phenotype. One field of research, transcriptomics, has revolutionized this central biological concept. Discovery of the abundance and complexity of RNA dynamics and function dramatically changed our understanding about the role of RNA, apart from encoding proteins, and challenged gene-centric approach to explain the function of genome.[2] Transcriptome is a collective term describing all RNAs produced by a single cell, by a population of cells or tissue.[3] Recent progress in analytical technologies has unveiled the complexity of the regulation of the transcriptome. The transcriptome is the primary product of the genome and therefore analysis of the transcriptome provides primary information for functional genomics.

The human genome

One of original and the most remarkable results of the human genome project was the discovery that only 1.2%

of the human genome encodes proteins and was therefore considered as functional and meaningful.[4] This is also reflected in the early search for genetic variation associated with a specific disease focusing on DNA sequences solely in exons. The number of genes, protein-coding elements, was discovered to be around 30,000, a much smaller number than predicted and similar to that found in several other species.[4] Based on these findings, the rest of the genome was initially termed as junk DNA. However, additional studies have identified that most of the DNA has function, not only for genome structure and packaging but also to form the complexity of the molecular networks underpinning the diversity of cell function. Early studies, after the identification of individual chromosome sequences, indicated that genomic sequences were transcribed at least as much as an order of magnitude more than accounted for by the predicted gene models.[2] Similarly, the term "pseudogene" that implies that is not a real gene and considered as a remnant of evolution or "genomic fossil."[5] It is now demonstrated that most of the pseudogenes are transcribed and translated into proteins challenging that definition of "pseudogene."[6]Cap-analysis gene expression (CAGE) technology enabled the identification of at least 180,000 transcripts in the mammalian genome, and it appeared that the majority of the genome is transcribed.[7] At least 60% of the genome has been described as a transcriptional forest, where transcription is performed from both strands of the same DNA region without gaps.[7] The most remarkable project in this field is known as an Encyclopedia of DNA Elements or ENCODE for short. Based on ENCODE findings, at least 80% of genome is actively transcribed, and this number is considered to be conservative.[8] Interpretation of such data requires changes in our view to the functional regulation of genome and that is a prerequisite for successful clinical translation of the genomics.

Transcriptome, transcriptomics, and transcriptome profiling

Transcriptome is a collection of the RNAs (transcripts) that single cell or tissue can produce, and it contains all types of RNAs.[9] Transcriptomics is the study of the transcriptome; analyzing RNA and its different subcategories (mRNA, micro-RNA, non-coding RNA, etc.) to identify changes in expression and its functional impact. Although transcriptomics focuses on content and transcript expression levels, it also includes the analysis of transcriptional regulation. The transcriptome can be studied by different methods; however, the most common options are genechips (to measure gene expression on microarray platform) and RNA sequencing (RNA-seq).[9] Gene expression arrays initially focused solely on polyA purified RNA that encode proteins. Moreover, genechips also suffer from the requirement to be pre-designed, i.e. the content on the array is based on our pre-existing knowledge of predominantly exons that can be easily identified in genome sequence data.[10,11] Therefore, genechips give us a snapshot of the transcriptional changes of mRNA, but this snapshot is rather limited. More recent arrays (transcript based and tiling arrays) can give very comprehensive information about the transcriptional changes, nevertheless the genechips are inherently bound to pre-existing knowledge and do not provide information about the sequences of the transcripts.[10,12] Only a few genechip versions are capable of identifying alternative splicing and specialized chip design is required to analyze such as micro-RNAs.[12] But the sequence information is lost in results files, and this is where the RNA-sequencing has clear advantage allowing for more detailed analysis to detect alternative splicing, intron retention, and other events reflecting alterations in transcriptome regulation and the other classes of RNA. Therefore, RNA-sequencing has become the main technology for transcriptome analysis.[9,13]

Sources of the transcriptome

Gene expression is both tissue specific and stimulus inducible; therefore, a key question for transcriptome analysis is the source of the tissue or cell type for analysis. The most common and easiest to justify is the primary tissue that is affected by pathological processes. This is based on the assumption that we know what tissue is affected, and we have some preliminary understanding what the timeline and mechanisms of the pathological changes are. However, this assumption can be deceiving. For example, with central nervous system disorders, it is difficult to determine which region or cell type is involved and also whether the pathological hallmarks of the disease were initiated by dysfunction in another brain regions, or periphery, many years before. As brain tissue is only accessible as postmortem tissue, the changes in the transcriptome could arise from selective alteration of gene expression by the postmortem time rather than in response to living with a chronic age-dependent disease occurring over a long time period.[14–16] In case of neurodegenerative diseases, this may mean that we miss the molecular pathological changes that initiate the degenerative process. The same is similar for other chronic age-dependent disease such as arthritis or heart disease. The cells that are targeted by primary pathology are often dead or have a significantly altered phenotype from those that represent the key pathological transitions. Some of the problems of addressing transcriptomics in the central nervous system are outlined below.

Firstly, recognized issues with the use of biobanked tissue samples that would affect transcriptomics include the heterogeneity of the samples, reliability of the diagnoses and variability in the quality control measures.[17] The most drastic example to illustrate reliability challenges comes from the biobank having 12,000 samples available for research and only 18 of them with the suitable information and quality by the end.[16] While the analysis of postmortem brain samples is still valid and informative from a research point of view, the impact of these studies to improve our understanding about neurodegenerative disease needs addressed in a broader context.[16] It is difficult to infer causative changes from the single time point that is based on the analysis of the tissues where the pathogenic processes are completed.

Secondly, subjects may have used drugs for a long time and depending on the course of the disease the treatment

schedules can be quite different between patients.[17] Moreover, it is quite realistic to assume that the subjects have had comorbidities and taken drugs for those symptoms as well. Drugs for heart disease and hypertension and statins are quite common in the aged population, and therefore, analysis of the postmortem samples should most certainly take into account the drug history and comorbidities as confounders. This is something we do not see very often in studies using postmortem tissue samples.

Thirdly, we need to consider what regions of the tissue are to be analyzed. Again, in the case of the brain, regional changes in gene expression can be enormous.[17] It is a complex tissue and choosing the right regions for comparison is often the most important decision for the analysis. For example, in the case of targeted mutation mouse models generated by homologous recombination the changes in the transcriptome of the brain are regionally very different.[18–20] Targeted mutant mouse lines allow exclusion of all confounding factors and careful matching of the study subjects for the genetically engineered mutations. However, even after the perfect matching for confounders, the deletion of the single gene induced enormously different changes in transcriptome in the different regions of the brain.[20] Only the lack of the expression of the deleted gene was the similar result between the different brain regions.[20] In addition to the regional difference in the brain tissue, genomic locus of the gene has also to be considered. We have analyzed the transcriptome of the Wolfram syndrome mutant mice with the deletion of the Wfs1 gene and identified significant confounding effect from the genomic locus of the targeted gene.[18] This locus-specific or genomic context effect means that even a single-gene targeting or deletion can induce the complex changes in the transcriptome that are not caused by the function of the gene, but by its location. Mouse models enable controlling for gender, age, and environmental differences, providing the ideal study design conditions, but cannot avoid genomic background effect, "congenic footprint."[21] This effect needs to be taken into account and with appropriate adjustment the functionally meaningful differences can be identified.[22] All this illustrates how diverse the transcriptome is in different brain regions, and therefore, it is challenging to design studies with multiple brain regions involved as it is not trivial to differentiate between the normal regional and pathologically relevant differences. In summary, by analyzing postmortem brains, we struggle to obtain the relevant information about the mechanisms of the disease, and this information does not always help us to design better diagnostic tools or drugs.

However, analysis of the diseased tissues is important when it is possible during the pathogenesis of the disease. Repeated sampling during the course of the disease allows us to use the time-dependent causative interaction models. Longitudinal studies are therefore the best way to follow disease progression but severely limit the choice of tissue or component that can be measured to such as blood, skin, urine, and microbiome. This also enables the monitoring of changes in the transcriptome during treatment and to compare different therapeutic options.[23] In more limited cases, surgical removal of tissue during medical procedures is another option to access samples for transcriptomic analysis. The latter option is the most common for oncological samples and is potentially applicable for any surgically treated conditions. If we plan to perform longitudinal transcriptome analysis with samples from different timepoints, then almost the only viable option is blood sampling. Skin sampling can also be alternative for some cases and diagnoses. We have shown that skin and blood are useful alternatives even for neurodegenerative diseases like Parkinson's disease.[24–26] Both blood and skin showed clear transcriptome differences in the case–control design, and these tissues could be used for the diagnosis or monitoring the progression of the disease. Similarly, urine can be used as a source for transcriptome analysis.[27,28] However, as usually the cellular content in urine is low, the RNA level is also low and that reduces potential of urine or other body fluids as a source for transcriptomics.[28]

Whole blood versus PBMC transcriptome

Blood is a useful and easy to access surrogate tissue for transcriptome analysis, but the use of blood requires a few basic decisions. For example, it is possible to analyze whole blood or a particular fraction of blood cells. Peripheral blood mononuclear cell (PBMC) separation has been one very popular method to isolate cells from the blood and to prepare them for RNA analysis. However, the PBMC fraction contains only lymphocytes and monocytes, while all granulocytes like basophils, eosinophils, and neutrophils are depleted. From all white cell count, neutrophils constitute 55% to 75% indicating that using of PBMC for transcriptome analysis would not give the full picture.[29] Isolation of PBMCs covers only 20% to 50% of the cellular heterogeneity of the blood. Moreover, PBMC separation itself is a procedure that adds an extra uncontrollable variation to the analysis, and this should be avoided. Several studies have shown significant differences between the transcriptome profiles between PBMC and whole blood.[30] It is reported that over 2000 genes were differentially expressed with more than two-fold difference between PBMC and whole blood from the same individual at same time.[31] Therefore, for transcriptome analysis, the whole blood RNA samples have a substantial advantage over PBMC or other fractionation.

Preanalytical considerations

Due to the complexity and the volume of the transcriptomics data, preanalytical conditions have significant impact on the outcome of the analysis. The inadvertent variations can be introduced with the sampling of the tissue, during the storage and transportation or by the differences in the extraction methods. In addition, as addressed in previous sections, the sources for RNA can be variable ranging from blood and other body fluids to the tissue biopsies, cellular smears, and to single-cell sorting. All these different approaches require standardized protocols to ensure reproducibility and high quality of the analysis. The testing and guidelines on how to prepare and purify different clinical samples are vital for the further implementation of the

transcriptomic analysis in clinical practice. RNA extraction can be notoriously complicated with variable options available that all can lead to different results.[32] Similarly, storage conditions have been shown to impact the quality of RNA and snap-frozen samples detect significantly more genes than formalin-fixed paraffin-embedded (FFPE) samples.[33] This effect was not dependent on the time to fixation. Interestingly, miRNA expression was not affected by the fixation method, and it was comparable between frozen or FFPE samples.[33] In addition, purification of the liquid biopsy samples requires an extra effort and a complex workflow.[34] As RNA can be purified from different samples, validation studies are required to develop standardized protocols that would enable robust and reproducible analysis of transcriptome for various clinical conditions.

Practical utility of transcriptome analysis

The transcriptome is a snapshot of molecular events in the cell reflecting the functional activity of the genome at a given moment of time and requires a combination of analytical tools to describe these molecular changes. Currently, the majority of genomic tools used in clinical genomics only consider targeted DNA sequencing and not the transcriptome. However, there are several examples of how transcriptomic information improves the precision of the genomic analysis.

The early studies to analyze transcriptomics used variable differential cloning technologies based on cDNA library preparation and comparative analysis.[35] One of these methods, cDNA representational difference analysis (cDNA-RDA), was used to identify differential expression in pancreatic cancer.[36]cDNA-RDA was proven to be a highly efficient and reproducible method that has been used in various models and organisms.[37,38] While the method itself was laborious and difficult to use for larger sample numbers, it clearly had its advantage as a hypothesis-free approach to observe transcriptional changes.[39] As the method did not require specific equipment or expensive preparations like gene microchips, the method gained popularity and was applied to study variable pathologies or physiological responses.[40] At the same time, cDNA microarray technology was also developing and provided various in-house products. These microarrays were based on the cDNA clone collections, their amplification, and printing (spotting) on to glass slides.[41] This technology required substantial infrastructure to run and it was not widely accessible. Nevertheless, initial studies demonstrated the suitability of this technology in pathology and particularly in cancer diagnostics where tumour material is available. These studies indicated that breast cancers can be classified by their gene expression patterns into subtypes that were not identifiable with histological methods alone.[42] The gene expression pattern was not only helpful to identify the molecular subtypes of the breast cancers but also to predict the clinical course and outcomes of breast cancer.[43] These early reports fuelled a myriad of similar studies to determine the transcriptional pattern of other tumors to identify potential diagnostic or prognostic biomarkers. Gene microarrays became standardized for transcriptional studies. The main advantage was the high-throughput analysis of the cDNA libraries, and as the technology was scalable, it was possible to increase sample sizes and the power of studies. However, gene arrays still suffered from a biased capture of targets which as stated previously were based on exon data or a limited number of non-coding RNAs. The latter was partially resolved when next generation sequencing (NGS) technologies become easily accessible to enable parallel whole genome sequencing (WGS) and RNA-sequencing (RNA-seq). Clinical genetics analysis rapidly expanded from exome sequences to a complete RNA analysis. RNA-seq is the first technology that enabled complete transcriptome analysis covering all different types of RNA subclasses with complete sequence information and enables detection of complex profiles from various pathologies.[9]

Several examples support the value and the utility of transcriptomics in the complex analysis of clinical samples for association to disease. We have analyzed the transcriptional profiles of osteosarcoma samples from fresh tumors in a paired study design and identified several new candidates involved in the development of osteosarcoma.[23] Moreover, with similar technology, we were able to analyze archived FFPE samples that gave us the possibility to evaluate the effect of chemotherapy on the transcriptional profile. The same data-set provided data regarding repetitive elements that were differentially expressed in the malignancy.[44] Repetitive elements can only be efficiently analyzed using the RNA-seq technology rather than genechips.

Transcriptome analysis can stratify patients who would otherwise be grouped as the same disease and this enables biomarker-driven clinical trials to improve their efficacy. Several meta-analyses have shown substantial improvement in study outcomes by using the biomarker-driven stratification in the study designs.[45] Personalized medicine approaches involving biomarkers in study design improved response rates from 5 to 30 percent demonstrating the improvement that can be achieved by using a genomics driven approach.[45] For example, the Winther trial based on 303 patients utilized genomic-matching to personalize their cancer therapy.[46] The study had two arms: one was based only on DNA data and the other only on RNA data. This trial introduced several innovative paradigm shifts showing an improved therapeutic response with the integration of transcriptomic profiling. Most importantly, the transcriptomic arm identified the most suitable solutions for the patients with various solid tumors prospectively from the large database of therapies.[46] This trial considered patient therapy options at an individual level based on the features of person's tumor and not on the results obtained from the aggregation of trials on large patient populations. Therapeutic guidance based only on the transcriptomic data resulted in the stabilizing of disease in 30% of patients.[46] While not statistically superior from the DNA-only approach (26%), it is was a remarkable success considering that the study subjects all had advanced cancers with several previous therapies that were unsuccessful. Transcriptomic-guided therapy was considered because the DNA analysis alone does not often reveal actionable variants or mutations and RNA analysis could

indicate the functional consequences. RNA-sequencing served here as an additional analytical tool to describe the functional changes in cancer that was in turn used in the therapeutic decision pipeline.

NGS technologies have also changed the ways we analyze Mendelian diseases and made whole-exome sequencing (WES) or WGS accessible to identify disease-causing variants. However, the success rate for detecting causal changes ranges only from 20% to 30%.[47] In a recent study, the use of RNA-seq analysis yielded diagnostic rate of 35% on previously unsolved cases by WGS analysis indicating a marked improvement.[48] The main advantage of RNA-seq is its ability to detect aberrant splicing or disruptive changes in the transcriptional regulation that are not detectable with WGS or WES.[48] This is the evidence to support the power of RNA-seq analysis also for Mendelian diseases and shows its clinical applicability in this space.

A recent example for the applicability of transcriptome analysis or RNA-based diagnostics can be found from the COVID-19 pandemic caused by the RNA-virus SARS-CoV2. The virus is only 29,900 bp long and contains 10 genes with gene 5 and 7 being functionally bicistronic.[49] Infection is based on the infectious transcriptome and can be viewed as a transcriptome infection. Maybe the efficient therapy for viral infections lies in the targeting of the transcriptome to affect their transcriptional capacity. Transcriptome-based therapies are already available for human diseases like Duchenne Muscular Dystrophy or amyloidosis showing the potential of the transcriptome-based therapeutics.[50–52]Transcriptome-based therapies offer a real systematic opportunity for personalized medicine, and it requires complex transcriptome analysis as input.[53] This therapeutic approach can turn the information in the transcriptomics into therapeutic options.

Conclusions

Transcriptomics is currently a rapidly evolving field with new data to either stand alone or integrate with other clinical information to expand and modify the future of health care. While current applications are mostly limited to experimental projects, a growing number of studies indicate the practical utility of transcriptomics for diagnostics, genomics-driven trial design, and personalized drug development. Larger clinical validation of such experimental hypothesis will allow for accepted clinical usage, indeed blood samples can be taken in general practice and sent off for analysis and interpretation centrally before transmission to the clinician. Transcriptomics has revealed the vast complexity of the transcriptome, and we are just beginning to understand the principles of how this translates to function, pathophysiology, and therapeutic opportunities.

Authors' contributions: GK, ALP, VJB, JPQ, and SK conceived the idea, performed literature search, drafted manuscript, and worked with the final version. All authors participated equally.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: ALP and SK are funded by MSWA, The Michael J. Fox Foundation, Shake It Up Australia, and The Perron Institute.

ORCID iD

Sulev Koks https://orcid.org/0000-0001-6087-6643

REFERENCES

1. Mattick JS. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* 2003;**25**:930–9
2. Kapranov P, Cawley SE, Drenkow J, Bekiranov S, Strausberg RL, Fodor SP, Gingeras TR. Large-scale transcriptional activity in chromosomes 21 and 22. *Science (New York, NY)* 2002;**296**:916–9
3. Velculescu VE, Zhang L, Zhou W, Vogelstein J, Basrai MA, Bassett DE Jr, Hieter P, Vogelstein B, Kinzler KW. Characterization of the yeast transcriptome. *Cell* 1997;**88**:243–51
4. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D,Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL,Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng J-F, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen H-C, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JGR, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AFA, Stupka E, Szustakowki J,Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang S-P, Yeh R-F, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Myers RM,

Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Patrinos A, Morgan MJ, International Human Genome Sequencing C, Whitehead Institute for Biomedical Research CfGR, The Sanger C, Washington University Genome Sequencing C, Institute UDJG, Baylor College of Medicine Human Genome Sequencing C, Center RGS, Genoscope, Cnrs UMR, Department of Genome Analysis IoMB, Center GTCS, Beijing Genomics Institute/Human Genome C, Multimegabase Sequencing Center TIfSB, Stanford Genome Technology C, University of Oklahoma's Advanced Center for Genome T, Max Planck Institute for Molecular G, Cold Spring Harbor Laboratory LAHGC, Biotechnology GBGRCf, *Genome Analysis G,Scientific management: National Human Genome Research Institute USNIoH, Stanford Human Genome C, University of Washington Genome C, Department of Molecular Biology KUSoM, University of Texas Southwestern Medical Center at D, Office of Science USDoE, The Wellcome T. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**:860–921

5. Tutar Y. Pseudogenes. *Comp Funct Genom* 2012;**2012**:424526
6. Kim MS, Pinto SM, Getnet D, Nirujogi RS, Manda SS, Chaerkady R, Madugundu AK, Kelkar DS, Isserlin R, Jain S, Thomas JK, Muthusamy B, Leal-Rojas P, Kumar P, Sahasrabuddhe NA, Balakrishnan L, Advani J, George B, Renuse S, Selvan LD, Patil AH, Nanjappa V, Radhakrishnan A, Prasad S, Subbannayya T, Raju R, Kumar M, Sreenivasamurthy SK, Marimuthu A,Sathe GJ, Chavan S, Datta KK, Subbannayya Y, Sahu A, Yelamanchi SD, Jayaram S, Rajagopalan P, Sharma J, Murthy KR, Syed N, Goel R, Khan AA, Ahmad S, Dey G, Mudgal K, Chatterjee A, Huang TC, Zhong J, Wu X, Shaw PG, Freed D, Zahari MS, Mukherjee KK, Shankar S, Mahadevan A, Lam H, Mitchell CJ, Shankar SK, Satishchandra P, Schroeder JT, Sirdeshmukh R, Maitra A, Leach SD, Drake CG, Halushka MK, Prasad TS, Hruban RH, Kerr CL, Bader GD, Iacobuzio-Donahue CA, Gowda H, Pandey A. A draft map of the human proteome. *Nature* 2014;**509**:575–81
7. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest AR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, Allen JE, Ambesi-Impiombato A, Apweiler R, Aturaliya RN, Bailey TL, Bansal M, Baxter L, Beisel KW, Bersano T, Bono H, Chalk AM, Chiu KP, Choudhary V, Christoffels A, Clutterbuck DR, Crowe ML, Dalla E, Dalrymple BP, de Bono B, Della Gatta G, di Bernardo D, Down T, Engstrom P, Fagiolini M, Faulkner G, Fletcher CF, Fukushima T, Furuno M, Futaki S, Gariboldi M, Georgii-Hemming P, Gingeras TR, Gojobori T, Green RE, Gustincich S, Harbers M, Hayashi Y, Hensch TK, Hirokawa N, Hill D, Huminiecki L, Iacono M, Ikeo K, Iwama A, Ishikawa T, Jakt M, Kanapin A, Katoh M, Kawasawa Y, Kelso J, Kitamura H, Kitano H, Kollias G, Krishnan SP, Kruger A, Kummerfeld SK, Kurochkin IV, Lareau LF, Lazarevic D, Lipovich L, Liu J, Liuni S, McWilliam S, Madan Babu M, Madera M, Marchionni L, Matsuda H, Matsuzawa S, Miki H, Mignone F, Miyake S, Morris K, Mottagui-Tabar S, Mulder N, Nakano N, Nakauchi H, Ng P, Nilsson R, Nishiguchi S, Nishikawa S, Nori F, Ohara O, Okazaki Y, Orlando V, Pang KC, Pavan WJ, Pavesi G, Pesole G, Petrovsky N, Piazza S, Reed J, Reid JF, Ring BZ, Ringwald M, Rost B, Ruan Y, Salzberg SL, Sandelin A, Schneider C, Schonbach C, Sekiguchi K,Semple CA, Seno S, Sessa L, Sheng Y, Shibata Y, Shimada H, Shimada K, Silva D, Sinclair B, Sperling S, Stupka E, Sugiura K, Sultana R, Takenaka Y, Taki K, Tammoja K, Tan SL, Tang S, Taylor MS, Tegner J, Teichmann SA, Ueda HR, van Nimwegen E, Verardo R, Wei CL, Yagi K, Yamanishi H, Zabarovsky E, Zhu S, Zimmer A, Hide W, Bult C, Grimmond SM, Teasdale RD, Liu ET, Brusic V, Quackenbush J, Wahlestedt C, Mattick JS, Hume DA, Kai C, Sasaki D, Tomaru Y, Fukuda S, Kanamori-Katayama M, Suzuki M, Aoki J, Arakawa T, Iida J, Imamura K, Itoh M, Kato T, Kawaji H, Kawagashira N, Kawashima T, Kojima M, Kondo S, Konno H, Nakano K, Ninomiya N, Nishio T, Okada M, Plessy C, Shibata K, Shiraki T, Suzuki S, Tagami M, Waki K, Watahiki A, Okamura-Oho Y, Suzuki H, Kawai J, Hayashizaki Y. The transcriptional landscape of the mammalian genome. *Science (New York, NY)* 2005;**309**:1559–63
8. Pennisi E. Genomics ENCODE project writes eulogy for junk DNA. *Science (New York, NY)* 2012;**337**:1159–61
9. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 2009;**10**:57–63
10. Dufva M. Introduction to microarray technology. *Methods Mol Biol (Clifton, NJ)* 2009;**529**:1–22
11. Iida K, Nishimura I. Gene expression profiling by DNA microarray technology. *Crit Rev Oral Biol Med* 2002;**13**:35–50
12. Della Beffa C, Cordero F, Calogero RA. Dissecting an alternative splicing analysis workflow for GeneChip exon 1.0 ST affymetrix arrays. *BMC Genom* 2008;**9**:571
13. Stark R, Grzelak M, Hadfield J. RNA sequencing: the teenage years. *Nat Rev Genet* 2019;**20**:631–56
14. Hunter MC, Pozhitkov AE, Noble PA. Accurate predictions of postmortem interval using linear regression analyses of gene meter expression data. *Forensic Sci Int* 2017;**275**:90–101
15. Pardue S, Zimmerman AL, Morrison-Bogorad M. Selective postmortem degradation of inducible heat shock protein 70 (hsp70) mRNAs in rat brain. *Cell Mol Neurobiol* 1994;**14**:341–57
16. Beach TG. Alzheimer's disease and the "valley of death": not enough guidance from human brain tissue? *J Alzheimer's Dis* 2013;**33**: S219–33
17. Wu C, Bendriem RM, Garamszegi SP, Song L, Lee CT. RNA sequencing in post-mortem human brains of neuropsychiatric disorders. *Psychiatry Clin Neurosci* 2017;**71**:663–72
18. Koks S, Soomets U, Paya-Cano JL, Fernandes C, Luuk H, Plaas M, Terasmaa A, Tillmann V, Noormets K, Vasar E, Schalkwyk LC. Wfs1 gene deletion causes growth retardation in mice and interferes with the growth hormone pathway. *Physiol Genom* 2009;**37**:249–59
19. Koks S, Soomets U, Plaas M, Terasmaa A, Noormets K, Tillmann V, Vasar E, Fernandes C, Schalkwyk LC. Hypothalamic gene expression profile indicates a reduction in G protein signaling in the Wfs1 mutant mice. *Physiol Genom* 2011;**43**:1351–8
20. Ivask M, Pajusalu S, Reimann E, Koks S. Hippocampus and hypothalamus RNA-sequencing of WFS1-deficient mice. *Neuroscience* 2018;**374**:91–103
21. Schalkwyk LC, Fernandes C, Nash MW, Kurrikoff K, Vasar E, Koks S. Interpretation of knockout experiments: the congenic footprint. *Genes Brain Behav* 2007;**6**:299–303
22. Koks S, Fernandes C, Kurrikoff K, Vasar E, Schalkwyk LC. Gene expression profiling reveals upregulation of Tlr4 receptors in Cckb receptor deficient mice. *Behav Brain Res* 2008;**188**:62–70
23. Ho XD, Phung P, Q Le V, H Nguyen V, Reimann E, Prans E, Koks G, Maasalu K, Le NT, H Trinh L, G Nguyen H, Martson A, Koks S. Whole transcriptome analysis identifies differentially regulated networks between osteosarcoma and normal bone samples. *Exp Biol Med (Maywood)* 2017;**242**:1802–11
24. Planken A, Kurvits L, Reimann E, Kadastik-Eerme L, Kingo K, Koks S, Taba P. Looking beyond the brain to improve the pathogenic understanding of Parkinson's disease: implications of whole transcriptome profiling of patients' skin. *BMC Neurol* 2017;**17**:6
25. Kurvits L, Reimann E, Kadastik-Eerme L, Truu L, Kingo K, Erm T, Koks S, Taba P, Planken A. Serum amyloid alpha is downregulated in peripheral tissues of Parkinson's disease patients. *Front Neurosci* 2019;**13**:13
26. Billingsley KJ, Lattekivi F, Planken A, Reimann E, Kurvits L, Kadastik-Eerme L, Kasterpalu KM, Bubb VJ, Quinn JP, Koks S, Taba P. Analysis of repetitive element expression in the blood and skin of patients with Parkinson's disease identifies differential expression of satellite elements. *Sci Rep* 2019;**9**:4369
27. Galichon P, Xu-Dubois YC, Buob D, Tinel C, Anglicheau D, Benbouzid S, Dahan K, Ouali N, Hertig A, Brocheriou I, Rondeau E. Urinary transcriptomics reveals patterns associated with subclinical injury of the renal allograft. *Biomark Med* 2018;**12**:427–38
28. Sole C, Goicoechea I, Goni A, Schramm M, Armesto M, Arestin M, Manterola L, Tellaetxe M, Alberdi A, Nogueira L, Roumiguie M, Lopez JI, Sanz Jaka JP, Urruticoechea A, Vergara I, Loizaga-Iriarte A, Unda M, Carracedo A, Malavaud B, Lawrie CH. The urinary transcriptome as a source of biomarkers for prostate cancer. *Cancers* 2020;**12**:513
29. Rhoades R, Bell DR. *Medical physiology. Principles for clinical medicine.* Online access with subscription: LWW health library (integrated basic

science collection). 4th ed. Philadelphia: Wolters Kluwer Health/ Lippincott Williams & Wilkins, 2013.
30. He D, Yang CX, Sahin B, Singh A, Shannon CP, Oliveria JP, Gauvreau GM, Tebbutt SJ. Whole blood vs PBMC: compartmental differences in gene expression profiling exemplified in asthma. *Allergy Asthma Clin Immunol* 2019;**15**:67
31. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci Usa* 2003;**100**:1896–901
32. Sourvinou IS, Markou A, Lianidou ES. Quantification of circulating miRNAs in plasma: effect of preanalytical and analytical parameters on their isolation and stability. *J Mol Diagn* 2013;**15**:827–34
33. Jones W, Greytak S, Odeh H, Guan P, Powers J, Bavarva J, Moore HM. Deleterious effects of formalin-fixation and delays to fixation on RNA and miRNA-Seq profiles. *Sci Rep* 2019;**9**:6980
34. Buschmann D, Haberberger A, Kirchner B, Spornraft M, Riedmaier I, Schelling G, Pfaffl MW. Toward reliable biomarker signatures in the age of liquid biopsies – how to standardize the small RNA-Seq workflow. *Nucleic Acids Res* 2016;**44**:5995–6018
35. Carulli JP, Artinger M, Swain PM, Root CD, Chee L, Tulig C, Guerin J, Osborne M, Stein G, Lian J, Lomedico PT. High throughput analysis of differential gene expression. *J Cell Biochem Suppl* 1998;**30–31**:286–96
36. Gress TM, Wallrapp C, Frohme M, Muller-Pillasch F, Lacher U, Friess H, Buchler M, Adler G, Hoheisel JD. Identification of genes with specific expression in pancreatic cancer by cDNA representational difference analysis. *Genes Chromosom Cancer* 1997;**19**:97–103
37. Hubank M, Schatz DG. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res* 1994;**22**:5640–8
38. Koks S, Luuk H, Nelovkov A, Areda T, Vasar E. A screen for genes induced in the amygdaloid area during cat odor exposure. *Genes Brain Behav* 2004;**3**:80–9
39. Bowler LD. Representational difference analysis of cDNA. *Methods Mol Med* 2004;**94**:49–66
40. Nelovkov A, Philips MA, Koks S, Vasar E. Rats with low exploratory activity in the elevated plus-maze have the increased expression of limbic system-associated membrane protein gene in the periaqueductal grey. *Neurosci Lett* 2003;**352**:179–82
41. Smith TP, Grosse WM, Freking BA, Roberts AJ, Stone RT, Casas E, Wray JE, White J, Cho J, Fahrenkrug SC, Bennett GL, Heaton MP, Laegreid WW, Rohrer GA, Chitko-McKown CG, Pertea G, Holt I, Karamycheva S, Liang F, Quackenbush J, Keele JW. Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle. *Genome Res* 2001;**11**:626–30
42. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. Molecular portraits of human breast tumours. *Nature* 2000;**406**:747–52
43. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, Borresen-Dale AL. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;**98**:10869–74
44. Ho XD, Nguyen HG, Trinh LH, Reimann E, Prans E, Koks G, Maasalu K, Le VQ, Nguyen VH, Le NTN, Phung P, Martson A, Lattekivi F, Koks S. Analysis of the expression of repetitive DNA elements in osteosarcoma. *Front Genet* 2017;**8**:193
45. Schwaederle M, Zhao M, Lee JJ, Lazar V, Leyland-Jones B, Schilsky RL, Mendelsohn J, Kurzrock R. Association of biomarker-based treatment strategies with response rates and progression-free survival in refractory malignant neoplasms: a meta-analysis. *JAMA Oncol* 2016;**2**:1452–59
46. Rodon J, Soria JC, Berger R, Miller WH, Rubin E, Kugel A, Tsimberidou A, Saintigny P, Ackerstein A, Brana I, Loriot Y, Afshar M, Miller V, Wunder F, Bresson C, Martini JF, Raynaud J, Mendelsohn J, Batist G, Onn A, Tabernero J, Schilsky RL, Lazar V, Lee JJ, Kurzrock R. Genomic and transcriptomic profiling expands precision cancer medicine: the WINTHER trial. *Nat Med* 2019;**25**:751–58
47. Taylor JC, Martin HC, Lise S, Broxholme J, Cazier J-B, Rimmer A, Kanapin A, Lunter G, Fiddy S, Allan C, Aricescu AR, Attar M, Babbs C, Becq J, Beeson D, Bento C, Bignell P, Blair E, Buckle VJ, Bull K, Cais O, Cario H, Chapel H, Copley RR, Cornall R, Craft J, Dahan K, Davenport EE, Dendrou C, Devuyst O, Fenwick AL, Flint J, Fugger L, Gilbert RD, Goriely A, Green A, Greger IH, Grocock R, Gruszczyk AV, Hastings R, Hatton E, Higgs D, Hill A, Holmes C, Howard M, Hughes L, Humburg P, Johnson D, Karpe F, Kingsbury Z, Kini U, Knight JC, Krohn J, Lamble S, Langman C, Lonie L, Luck J, McCarthy D, McGowan SJ, McMullin MF, Miller KA, Murray L, Németh AH, Nesbit MA, Nutt D, Ormondroyd E, Oturai AB, Pagnamenta A, Patel SY, Percy M, Petousi N, Piazza P, Piret SE, Polanco-Echeverry G, Popitsch N, Powrie F, Pugh C, Quek L, Robbins PA, Robson K, Russo A, Sahgal N, van Schouwenburg PA, Schuh A, Silverman E, Simmons A, Sørensen PS, Sweeney E, Taylor J, Thakker RV, Tomlinson I, Trebes A, Twigg SR, Uhlig HH, Vyas P, Vyse T, Wall SA, Watkins H, Whyte MP, Witty L, Wright B, Yau C, Buck D, Humphray S, Ratcliffe PJ, Bell JI, Wilkie AO, Bentley D, Donnelly P, McVean G. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nat Genet* 2015;**47**:717–26
48. Cummings BB, Marshall JL, Tukiainen T, Lek M, Donkervoort S, Foley AR, Bolduc V, Waddell LB, Sandaradura SA, O'Grady GL, Estrella E, Reddy HM, Zhao F, Weisburd B, Karczewski KJ, O'Donnell-Luria AH, Birnbaum D, Sarkozy A, Hu Y, Gonorazky H, Claeys K, Joshi H, Bournazos A, Oates EC, Ghaoui R, Davis MR, Laing NG, Topf A, Kang PB, Beggs AH, North KN, Straub V, Dowling JJ, Muntoni F, Clarke NF, Cooper ST, Bonnemann CG, MacArthur DG. Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci Transl Med* 2017;**9**:eaal5209
49. Irigoyen N, Firth AE, Jones JD, Chung BY, Siddell SG, Brierley I. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. *PLoS Pathog* 2016;**12**:e1005473
50. Benson MD, Dasgupta NR, Monia BP. Inotersen (transthyretin-specific antisense oligonucleotide) for treatment of transthyretin amyloidosis. *Neurodegener Dis Manag* 2019;**9**:25–30
51. Wein N, Vulin A, Findlay AR, Gumienny F, Huang N, Wilton SD, Flanigan KM. Efficient skipping of single exon duplications in DMD patient-derived cell lines using an antisense oligonucleotide approach. *J Neuromuscul Dis* 2017;**4**:199–207
52. Li D, Mastaglia FL, Fletcher S, Wilton SD. Precision medicine through antisense oligonucleotide-mediated exon skipping. *Trends Pharmacol Sci* 2018;**39**:982–94
53. Aung-Htut MT, McIntosh CS, Ham KA, Pitout IL, Flynn LL, Greer K, Fletcher S, Wilton SD. Systematic approach to developing splice modulating antisense oligonucleotides. *Int J Mol Sci* 2019;**20**:5030