

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

PARSE BIOSCIENCES, INC.,
Petitioner

v.

10X GENOMICS, INC.,
Patent Owner

Case IPR2023-00955
Patent 10,240,197

EXPERT DECLARATION OF JOHN QUACKENBUSH, PH.D

TABLE OF CONTENTS

I.	INTRODUCTION.....	2
II.	BACKGROUND AND QUALIFICATIONS.....	2
III.	MATERIALS RELIED UPON.....	8
IV.	LEGAL STANDARDS	8
	A. Obviousness.....	8
	B. One Of Ordinary Skill	11
V.	BACKGROUND ON THE STATE OF THE ART AND THE CHALLENGED CLAIMS	12
	A. The '197 Patent's Invention	12
	B. Overview Of Prior Art.....	21
	1. Linnarsson (EX1003).....	21
	2. McCloskey (EX1004)	38
	3. McCloskey II (EX1055)	42
VI.	Claim Construction.....	42
VII.	THE CHALLENGED CLAIMS ARE VALID	43
	A. Ground 1: Linnarsson In View of McCloskey Does Not Render Claims 1-4, 6-11, 20 or 23-25 Obvious.....	43
	1. Dr. Cooper Fails To Show Linnarsson Teaches “Sequencing The Tagged Polynucleotide”	44
	2. McCloskey Does Not Disclose A “First Tag” Sequence.....	50
	3. McCloskey's Batch-Stamp Lacks The Benefits Of Linnarsson's Cell-Tag For Single-Cell Analysis.....	55
	4. There Are Not Enough Unique Tags In McCloskey To Tag “Each Polynucleotide” In Linnarsson.....	57
	5. McCloskey Does Not Describe Single-Cell Analysis	63
	6. Dr. Cooper's Claim 1 Reasonable Expectation of Success Fails.....	65
	B. Ground 2: Linnarsson In View Of McCloskey In View Of Knowledge Of A POSA Does Not Render Claim 5 Obvious.....	65
	1. Dr. Cooper's Theory	65

2.	Dr. Cooper’s Motivation To Modify Dr. Cooper’s Ground 1 Fails	66
3.	Dr. Cooper Has Not Shown A Tag Can Be Successfully Added Via Ligation In Linnarsson While Still Practicing The Claimed Method	69
4.	Dr. Cooper’s Other Ligation Assertions Fail.....	76
5.	Dr. Cooper’s Ligation Theory Fails If Ligation Occurs After Amplification.....	83
C.	Ground 3: Linnarsson In View Of McCloskey In View Of McCloskey II Does Not Render Claim 13 Obvious	84
VIII.	CONCLUSION	85

I, Dr. John Quackenbush, hereby declare under penalty of perjury:

I. INTRODUCTION

1. I have personal knowledge of the facts and opinions set forth in this declaration, and if called upon to do so, I would testify competently thereto.

2. I have been retained by Groombridge, Wu, Baughman & Stone LLP, on behalf of Patent Owner 10x Genomics, Inc. (“Patent Owner” or “10x”) to offer opinions regarding certain issues relating to U.S. Patent No. 10,240,197 (the “’197 Patent”), assigned to 10x, as well as other documents presented to me by counsel for 10x, in connection with *Inter Partes* Review of U.S. Patent No. 10,240,197 under 35 U.S.C. §§ 311-319 and 37 C.F.R. § 42 (“Petition”).

3. I am being compensated at a rate of \$750/hour for time testifying and \$750/hour for my other time spent on this proceeding and reimbursed for expenses incurred as a result of my role as an expert. My compensation is not in any way contingent upon my performance, the result of this proceeding, or any of the issues involved therein. I have no financial interest in the outcome of this case.

II. BACKGROUND AND QUALIFICATIONS

4. I am the Henry Pickering Walcott Professor of Computational Biology and Bioinformatics and Chair of the Department of Biostatistics at the Harvard T.H. Chan School of Public Health. I am also Founding Director and co-Director of the MS Program in Computational Biology and Quantitative Genetics and Director of

the Big Data to Knowledge Training Program, co-Director of the Cancer Genomics Training Program, and the Faculty Director for the Harvard Quantitative Biomedical Research Center. In addition, I hold appointments as Professor of Biostatistics and Computational Biology at the Dana-Farber Cancer Institute and at the Channing Division of Network Medicine of Brigham and Women's Hospital.

5. In addition to my academic appointments, I co-founded Genospace LLC in 2012. Genospace produces software solutions to operationalize precision medicine through the integration of clinical and genomic data and its presentation to clinical and research users in an intuitive format. Genospace was purchased by Hospital Corporation of America in January of 2017.

6. In addition to my academic and research responsibilities, over the years I have taken significant regional and national research leadership responsibilities. I have served on the Scientific Advisory Board of the St. Jude's Children's Research Hospital and the Board of Scientific Counselors of the Environmental Protection Agency. I have also served on numerous scientific review panels for the National Institutes of Health, the National Science Foundation, the Department of Energy, the National Aeronautics and Space Administration, the Wellcome Trust, Genome British Columbia, Genome Canada, the National Sciences and Engineering Council of Canada, and the European Union. And I have served on numerous study panels

for the National Research Council, the National Academy of Sciences, and the Institute of Medicine.

7. I have presented more than 350 talks, courses, and workshops since joining the Harvard School of Public Health and Dana-Farber in 2005. I have also taught numerous courses and workshops on the analysis of genomic data and served as an organizer of the EMBO course on DNA microarray analysis for more than 5 years.

8. Prior to joining the faculty at Dana-Farber and Harvard, I was a Professor of Chemical Engineering at the University of Maryland from 2003 – 2005; an Investigator with The Institute for Genomic Research from 2002 – 2006; a Professor in the Department of Biochemistry at The George Washington University from 2000 – 2005; a Lecturer in The Department of Biostatistics at The Johns Hopkins University from 1998 – 2005; an Associate Investigator at The Institute for Genomic Research from 1997 – 2001; and an Assistant Investigator at The Institute for Genomic Research in 1997.

9. I received my Ph.D. in Theoretical Particle Physics at the University of California, Los Angeles in 1990. I also hold a Master of Science in Physics from University of California, Los Angeles and a Bachelor of Science in Physics from the California Institute of Technology.

10. I have studied, taught, practiced, and conducted research involving the use of genomic technologies, including DNA sequencing and DNA microarrays, and involving computational analysis of data related to genomic technologies. I have expertise in genomic technologies, computational analysis, and modeling of a plant, animal, and microbial systems with an emphasis on the study of human diseases, such as cancer.

11. In describing my work, I often use terms to refer to techniques and technologies, as well as specific application areas. During the course of my career, I have worked in genomics, functional genomics, epigenomics, bioinformatics, biostatistics, computational biology, systems biology, network biology, and network medicine. Although the boundaries between these various fields are often not well defined, one possible distinction is that some are directed toward data generation while others are directed toward developing and applying methods for data analysis and interpretation. For example, genomics, functional genomics, and epigenomics are generally associated with the design and conduct of experiments for generation of large-scale data on hundreds, thousands, or tens of thousands of genes, RNA transcripts, or epigenetic sites within the genome (or other alterations such as identifying regulatory interactions or structural changes in the genome). Fields including bioinformatics, biostatistics, computational biology, systems biology, network biology, and network medicine generally involve the development of new

methods for data analysis and application of those methods to the biological and biomedical data sets, including those generated using techniques such as genomics, functional genomics, and epigenomics. The beauty of all of these techniques is that their development and use is, largely although not entirely, independent of their application area. So methods developed in one area are often used in other areas.

12. I have worked in the analysis of genes in hundreds of animal, plant, unicellular, and parasitic species, including human, mouse, rat, cattle, sheep, brown bear, salmon, corn, tomato, rice, yeast, the causative agent of malaria, and countless bacteria and viruses. Despite this diversity of applications, my primary research emphasis has been in the study of human diseases, with the greatest body of my recent work in the analysis of human cancers and pulmonary diseases. However, I have also explored a wide range of other issues in health and disease, including Alzheimer's disease, Parkinson's disease, Huntington's disease, neonatal sepsis, and Down syndrome. My work has involved analysis of data generated from bulk tissue samples, small numbers of cells, single cells, and cell-free DNA.

13. I have also done work relevant to reproductive health, an area that became personally important to me as my wife and I, then in our forties, became enmeshed in the use of assisted reproductive technology. This gave me first-hand exposure to a wide range of techniques, including pre-implantation prenatal screening and the use and limitations of technologies directed toward prenatal

detection of genetic abnormalities including aneuploidies. Following the birth of my son in 2006 at Brigham and Women's Hospital in Boston, I also worked with, and mentored, physician scientists at Harvard, published a genomic analysis of diminished ovarian reserve, and presented the President's plenary lecture at the 63rd American Society for Reproductive Medicine Annual Meeting in 2007 in which I described the use of genomic technology in reproductive, prenatal, and neonatal health.

14. I have received numerous awards and honors including being recognized as a White House Open Science Champion of Change, an Australia Fellow, and a Distinguished International Advisor for the International Conference of BioInformatics and BioEngineering sponsored by the Institute of Electrical and Electronics Engineers. In 2022 I was elected to the National Academy of Medicine, one of three academies that make up the National Academies of Sciences, Engineering, and Medicine (the National Academies) in the United States.

15. I have published more than 320 peer reviewed scientific papers that have been cited more than 94,000 times. I have had continuous grant support for my research from the National Institutes of Health and other organizations for nearly thirty years. In 2018, I received an NCI Outstanding Investigator Career Award, a seven-year research grant recognizing a lifetime record of substantial contributions to the field.

16. Attached hereto as **Appendix A** is a true and correct copy of my Curriculum Vitae describing my background and experience and includes a list of my publications, presentations, and expert testimony experience.

III. MATERIALS RELIED UPON

17. In preparing this Declaration, I have relied upon the following: the '197 Patent (EX1001) and its File History (EX1005), the Declaration of Gregory Cooper, Ph.D. (EX1002), and EX1003, EX1004, EX1006, EX1007, EX1009, EX1013, EX1035, EX1055, EX2003, EX2007, EX2012, EX2014, EX2015, EX2016, EX2017, EX2018, EX2019, EX2020, the Rehearing Decision, and all of the materials cited in this Declaration.

IV. LEGAL STANDARDS

A. Obviousness

18. I understand that a claim is obvious pursuant to 35 U.S.C. § 103 if the differences between the claimed invention and the prior art are such that the claimed invention, as a whole, would have been obvious to persons having ordinary skill in the art (“persons of skill” or “POSITA”) to which the claimed invention pertains—before the earliest filing date of the claimed invention. I understand that a claim is valid unless the differences between the claimed invention and the prior art are such that the subject matter as a whole would have been obvious at the time of the

invention to persons of skill. Here, I have been asked to assume that the earliest filing date of the '197 patent is August 20, 2009.

19. I understand that the following factors are used to make an obviousness determination: (i) the scope and content of the prior art; (ii) the level of ordinary skill in the art; (iii) the differences between the claimed invention and the prior art; and (iv) other “objective indicia” (sometimes called “secondary considerations”) of non-obviousness.

20. I also understand that for a claim to be found obvious, each element must be disclosed in the prior art. I understand, however, that when support for obviousness is argued based on a combination of prior art references, an invention is not proved obvious merely by demonstrating that each of the elements in the claim was independently known in the prior art.

21. Rather, I understand that for obviousness, there must be a reason or motivation that would have prompted persons of skill in the relevant field to combine or modify the elements in the way the claimed invention does. I understand that using the claimed invention as a blueprint to piece together various elements in the prior art amounts to hindsight reasoning and is prohibited.

22. I also understand that, for a patent to be obvious, persons of skill must have had, in addition to a motivation to combine, a reasonable expectation of success in practicing the claimed invention based on the prior art.

23. I further understand that, in making a determination as to whether or not the claimed invention would have been obvious to persons of skill, the Board must consider certain objective “indicia” or factors, such as commercial success (e.g., the profitability of the claimed method or device practicing the claims), long-felt but unsolved need, teaching away (e.g., literature telling persons of skill not to do something that would be required by the claims), unexpected results, copying, and praise by others in the field. The presence of such factors is evidence of nonobviousness. I also understand that a connection, or nexus, must exist between the objective factor and the claimed invention.

24. I understand that, among other things, the Board has instituted an inter partes review of claims 1-26 of the '197 Patent. I understand that the instituted grounds are as follows: Ground 1 challenges claims 1-4, 6-11, 20, and 23-25 of the '197 Patent as unpatentable as obvious over the combination of Linnarsson (EX1003) in view of McCloskey (EX1004), Ground 2 challenges claims 5, 12, 21, 22, and 26 of the '197 Patent as unpatentable as obvious over the combination of Linnarsson (EX1003) in view of McCloskey (EX1004) and further in view of the knowledge of a POSA, and that Ground 3 challenges claims 13-19 as unpatentable as obvious over the combination of Linnarsson (EX1003) in view of McCloskey (EX1004) and further in view of McCloskey II (EX1055). I understand that for *inter*

partes review, invalidity must be shown under a preponderance of the evidence standard.

B. One Of Ordinary Skill

25. I have been asked to assume that persons of skill of the '197 Patent would have had Ph.D. in molecular biology, molecular genetics, biochemistry, physics, mechanical engineering, or a related discipline, with two years of academic, research, or industry experience in the field of genomic or transcriptome sequencing. I understand Dr. Cooper based his opinions on a different definition of a persons of skill, which is an individual with “a Masters degree and/or Ph.D. in Molecular Biology, Genetics, Chemistry, Engineering, or a related discipline and at least two years of post-Masters, postdoctoral, or industry experience. Alternatively, a POSA would have been someone with a Bachelor of Science degree in Molecular Biology, Genetics, Chemistry, Engineering, or a related discipline and at least five years of industry or academic experience (including any experimental work toward a graduate degree) relating to DNA sequencing technologies, amplification methods, and gene expression analysis” and “would have been familiar with associated tools, methods, and techniques including: (1) “polynucleotide tagging”; (2) “amplification”; (3) “use of tags to overcome amplification bias”; (4) “single cell analysis”; (5) “ligation and primer extension”; and (6) “sequencing.” EX1002

(Cooper Declaration), ¶¶53. My opinions would not change if I applied that definition of a persons of skill.

26. Given my education and professional experience described above and in my curriculum vitae, I would have qualified as a person of skill as of August 20, 2009 (under both my definition and Dr. Cooper's), and I have applied the knowledge of persons of skill in my analysis below. Well before August 20, 2009, my level of skill in the art was at least that of a person of skill. I am qualified to provide opinions concerning what persons of skill would have known and understood at that time, and my analysis and conclusions herein are from the perspective of persons of skill as of August 20, 2009. I understand that a person of skill is presumed to have knowledge of all relevant prior art, and therefore would have been familiar with each of the references cited herein and the full range of teachings they contain.

V. BACKGROUND ON THE STATE OF THE ART AND THE CHALLENGED CLAIMS

A. The '197 Patent's Invention

27. The '197 Patent is titled "Methods For Analyzing Nucleic Acids From Single Cells." EX1001, 1. I understand that the '197 Patent claims priority to an August 20, 2009 provisional, and I have used that date for assessing the prior art and Dr. Cooper's opinions. EX1001, 1; EX1002, ¶37. The '197 Patent's Challenged Claims relate to methods of counting nucleic acids from a plurality of single cells

that allow the user not only to differentiate among the different polynucleotides in an overall sample (including *e.g.*, mRNA), but also to track from which particular cell each polynucleotide originated. EX1001, 6:33-7:22, 15:7-50, claim 1 (30:18-48).

28. The patent describes using a multiplex identifier (“MID”) with two tag sequences: “a first tag sequence associated with the single cell from which the sample polynucleotide is derived” and “a second tag sequence distinguishing the sample polynucleotide from other sample polynucleotides derived from the same cell.” *Id.* Put differently, the claimed method employs both (1) cell-specific tags and (2) polynucleotide-specific tags. *Id.* Cells contain many types of polynucleotides, including DNA such as nuclear DNA and mitochondrial DNA (mtDNA), and RNA such as transfer RNA (tRNA), messenger RNA (mRNA), ribosomal RNA (rRNA), micro RNA (miRNA), and small nucleolar RNA (snoRNA), each of which can be analyzed using the claimed invention. EX1001, 14:8-14. According to Dr. Cooper, a single mammalian cell has between 50,000 and 300,000 mRNA molecules alone. EX1002, ¶255; EX1054, Abstract. As the ’197 patent notes, “In general, to accomplish such molecular tagging the number of distinct MID tags to be used should be many times greater than the number of molecules to be analyzed.” EX1001, 22:54-57. According to Dr. Cooper, tagging each molecule would require “at least 100 times more tags than sample

polynucleotides.” EX1002, ¶253. In other words, applying Dr. Cooper’s tag ratio, tagging each mRNA molecule in a mammalian cell requires **5,000,000 – 30,000,000 unique tags**. *Id.*; EX1054, Abstract. In my opinion, persons of skill would choose to err towards the high end of that range, since it is impossible to know in advance whether the cell of interest contains mRNA molecules at the low end of the expected range (50,000 by Dr. Cooper’s analysis) or at the high end (300,000).

29. Until inventions such as the ’197 Patent, most nucleotide sequencing was performed using polynucleotides (e.g., DNA or RNA) from a population of cells in a process called bulk analysis. EX1003, 2; EX1006, 2. Bulk analysis provided users an understanding of the polynucleotides found in the overall population of cells in the sample, but was unable to capture key differences among *individual* cells. EX1002, ¶73. For example, overexpression of the HER2 gene is associated with an aggressive form of breast cancer, but in the early stages of cancer a biopsy sample might contain only a small number of cancerous cells among many stromal and other cells. Bulk RNA sequencing analysis might well detect the expression of the HER2 gene, but would almost certainly not detect an overexpression of HER2 in the cancer cells, as the signal would be diluted as the detected bulk signal would be “averaged” *across the entire population of cells* in a sample. Thus, analyzing samples through a bulk analysis would not provide a critical piece of information. In contrast, single-cell RNA sequencing as enabled by the ’197 patent would allow not only

overexpression of HER2 in individual cells within a multi-cell sample to be identified, but also would provide a quantitation of the number of tumor cells in the sample. *See* EX1002, ¶¶73-74.

30. The specification of the '197 Patent explains that tagging a polynucleotide with the '197 Patent's disclosed first and second tag sequences before amplification, identifying the source cell and the individual polynucleotide respectively, provides a "correlation between a polynucleotide and its source" so that researchers can determine, for example, how many individual polynucleotides are in each single cell within a sample. EX1001, 6:38-51. The innovative methods claimed in the '197 Patent thus permitted users to do what had not previously been possible: to determine the quantity of *each polynucleotide present* in a cell *on a cell-by-cell basis* and to determine the quantity of each type of individual cell. EX1001, 6:33-7:22, claim 1 (30:18-48). Applying the inventive methods of the '197 Patent, the medical professional of the example above could identify individual cells with over-expressed HER2 genes and could then diagnose the patient with cancer at an early stage.

31. I disagree with certain of Dr. Cooper's characterizations of the specification of the '197 Patent. For example, Dr. Cooper states that the '197 tagging method is not new, but rather a prior art technique. EX1002, ¶29 (citing EX1001, 1:26-29). I disagree. The portion of the specification of the '197 Patent Dr. Cooper

cites does not describe dual tagging in the manner claimed. EX1002, ¶29 (citing EX1001, 1:26-29). Indeed, that portion, excerpted below, does not describe what tags are used, let alone “reflect[] that the tagging method is not the invention but a prior art technique.” EX1002, ¶29. Instead, the passage in the ’197 patent that Dr. Cooper cites clearly relates to a technique known as “multiplex” sequencing in which individual DNA samples (from distinct bulk sources) are labeled with sample-specific molecular barcodes (or “sample bar codes”) during library construction, multiple sample libraries are pooled and sequenced (referred to as “multiplex sequencing”) and the resulting data “demultiplexed” to group sample-specific sequences based on their sample bar codes.

We have previously described methods that enable tagging each of a population of fragmented genomes and then combining them together to create a ‘population library’ that can be processed and eventually sequenced as a mixture.
--

32. The Challenged Claims, as broken out by Dr. Cooper, are listed below:

[1A] A method of counting nucleic acids in a sample, the method comprising:

[1B] (a) providing a sample comprising a plurality of cells, wherein a cell of the plurality of cells comprises a plurality of sample polynucleotides;

[1C] (b) generating a plurality of tagged polynucleotides from the plurality of sample polynucleotides of said cell and a plurality of oligonucleotide tags,

[1D] wherein a tagged polynucleotide of the plurality of tagged polynucleotide comprises: (i) a sample sequence from a sample polynucleotide of the plurality of sample polynucleotides;

[1E] wherein a tagged polynucleotide of the plurality of tagged polynucleotide comprises: . . . (ii) a first tag sequence distinguishing said sample from sample polynucleotides from other cells; and

[1F] wherein a tagged polynucleotide of the plurality of tagged polynucleotide comprises: . . . (iii) a second tag sequence distinguishing said sample polynucleotide from other sample polynucleotides from said same cell;

[1G] (c) sequencing the tagged polynucleotide to determine the sample sequence, the first tag sequence, and the second tag sequence;

[1H] (d) using the first tag sequence and the second tag sequence to count a number of sample polynucleotides in said plurality of sample polynucleotides of said cell.

Claim 2: The method of claim 1, wherein the method further comprises amplifying the plurality of tagged polynucleotides prior to the sequencing step (c).

Claim 3: The method of claim 1, wherein the plurality of sample polynucleotides are selected from DNA and RNA.

Claim 4: The method of claim 3, wherein the plurality of sample polynucleotides comprise mRNA.

Claim 5: The method of claim 1, wherein the plurality of tagged polynucleotides is generated through at least one ligation reaction.

Claim 6: The method of claim 1, wherein an oligonucleotide tag of said plurality of oligonucleotide tags comprises said first tag sequence and said second tag sequence.

Claim 7: The method of claim 1, wherein an oligonucleotide tag of said plurality of oligonucleotide tags comprises a sequence that is configured to hybridize to said sample polynucleotide.

Claim 8: The method of claim 1, wherein the plurality of tagged polynucleotides is generated by (i) hybridizing an oligonucleotide tag of said plurality of oligonucleotide tags to said sample polynucleotide and (ii) extending said oligonucleotide tag or said sample polynucleotide or both.

Claim 9: The method of claim 1, wherein the plurality of tagged polynucleotides is generated through at least one linear amplification reaction.

Claim 10: The method of claim 1, wherein the plurality of tagged polynucleotides is generated through at least one reverse transcription reaction.

Claim 11: The method of claim 1, wherein the plurality of tagged polynucleotides is generated through at least one polymerase chain reaction (PCR).

Claim 12: The method of claim 1, wherein substantially every sample polynucleotide of said plurality of sample polynucleotides is associated with the same first tag sequence.

Claim 13: The method of claim 1, wherein at least 90 percent of said plurality of sample polynucleotides is associated with a unique second tag sequence.

Claim 14: The method of claim 1, wherein at least 95 percent of said plurality of sample polynucleotides is associated with a unique second tag sequence.

Claim 15: The method of claim 1, wherein at least 99 percent of said plurality of sample polynucleotides is associated with a unique second tag sequence.

Claim 16: The method of claim 1, wherein substantially every sample polynucleotide of said plurality of sample polynucleotides is associated with a unique second tag sequence.

Claim 17: The method of claim 1, wherein the number of different second tag sequences is larger than the number of sample polynucleotides.

Claim 18: The method of claim 17, wherein the number of different second tag sequences is at least ten times the number of sample polynucleotides.

Claim 19: The method of claim 18, wherein the number of different second tag sequences is at least one hundred times the number of sample polynucleotides.

Claim 20: The method of claim 1, wherein said plurality of sample polynucleotides are randomly associated with said plurality of oligonucleotide tags to generate said plurality of tagged polynucleotides.

Claim 21: The method of claim 1, wherein step (c) comprises hybridizing said tagged polynucleotide to a solid support.

Claim 22: The method of claim 20, wherein said solid support is a bead.

Claim 23: The method of claim 1, wherein step (d) comprises using said first tag sequence to distinguish (i) the number of sample polynucleotides

having said sample sequence from said cell from (ii) the number of sample polynucleotides having said sample sequence from other cells.

Claim 24: The method of claim 1, wherein step (d) comprises determining the number of different second tag sequences associated with said sample sequence, thereby estimating the number of sample polynucleotides having said sample sequence from said cell.

Claim 25: The method of claim 1, wherein said plurality of sample polynucleotides comprises substantially all mRNA molecules of said cell.

Claim 26: The method of claim 1, wherein said plurality of sample polynucleotides comprises a subset of polynucleotides of said cell having the same sequence.

B. Overview Of Prior Art

1. Linnarsson (EX1003)

33. EX1003 (“Linnarsson”) discloses a “single-cell tagged reverse transcription (STRT) method.” EX1003, 5:10-11, 23:1-4 (Example I disclosing STRT method), 31:3-5 (Example II disclosing STRT method); 37:31-32; *see generally* Pet.24-29. Linnarsson’s method involves preparing a cDNA library for sequencing by reverse-transcribing mRNA initially produced by transcription of

genes being expressed in the analyzed cells. EX1003, 1 (Abstract), 3:25-4:2, 12:4-8 (defining cDNA library as “a collection of cloned complementary DNA (cDNA) fragments”)¹; EX1002, ¶104. In a step of this process, “single cells are placed in separate wells of a 96-well plate and lysed....” EX1003, 4:8-13. Each lysed cell “contain[s] a plurality of mRNA molecules.” EX1003, 3:26-30; EX1054, Abstract; EX1002, ¶131. Following lysis, Linnarsson adds cDNA synthesis primers (“CDSs”) and/or template switching oligonucleotides (“TSOs”), which are then incorporated into cDNA during reverse transcription from mRNA. EX1003, 18:32-33. According to Linnarsson, a sequence in the CDSs and/or TSOs can be used to identify the cell from which the tagged polynucleotide originated. EX1003, 19:4-11.

¹ Persons of skill would have understood this to refer to fragments of the reverse-transcribed cDNA. This is supported by Linnarsson as a whole, which relies on sizing the reverse-transcribed polynucleotides such that they are suitable for sequencing. *Infra*, ¶¶45-47.

34. The CDSs are designed to anneal² to mRNA molecules. According to Dr. Cooper, “Linnarsson’s cDNA synthesis primer... includes a sequence of thymines...” EX1002, ¶105. Because virtually every eukaryotic mRNA molecule has a poly-A tail, the CDSs with a sequence of thymines, also referred to as an oligo-dT, can hybridize or anneal to virtually every mRNA molecule in a well. EX1003, 18:27-30; *infra* n.14.

35. The TSOs are designed to anneal to newly synthesized single-stranded cDNA from reverse transcribed mRNA. EX1003, 16:18-17:2; 18:19-26. Linnarsson’s TSOs include a poly-guanosine sequence. EX1003, 18:19-21. In the process of reverse-transcribing mRNA into cDNA, two to five cytosines are typically added to the 3’ end of the single-stranded cDNA by the reverse transcriptase. EX1003, 16:18-31. The poly-guanosine sequence in TSOs “‘anneals transiently’ to the series of cytosine nucleotides at the end of [a] newly-synthesized strand.” EX1003, 18:19-23. The TSO annealing to cDNA causes “the reverse transcriptase to switch template and continue to synthesis [sic] a strand

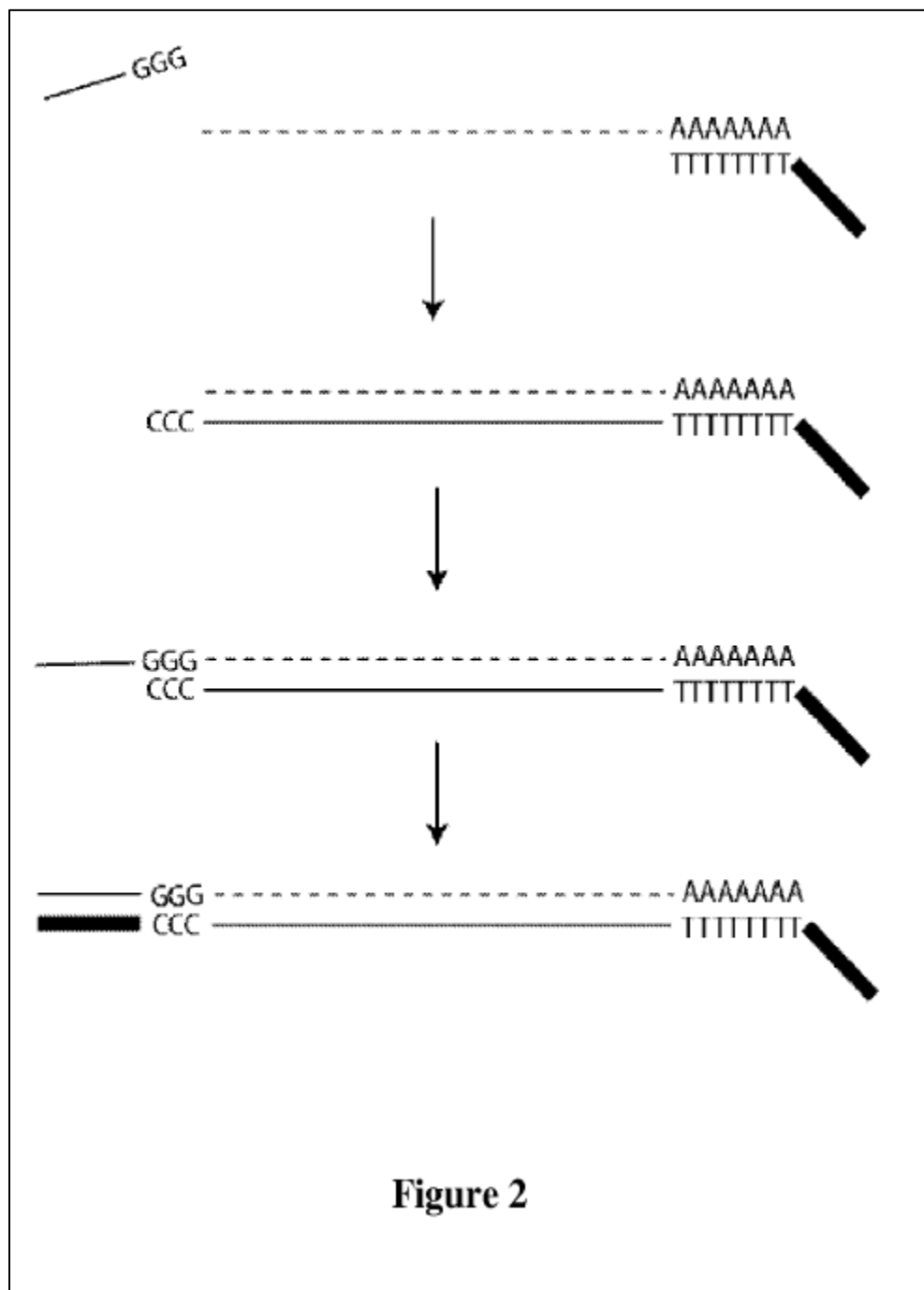
² Annealing is the joining of two complementary strands of nucleic acids through hydrogen bonding. EX1002, ¶¶56, 86; EX1003, 13:3-10; EX1009, 1 (“Denaturation separates the complementary strands of DNA held together in the duplex by hydrogen bonds.”).

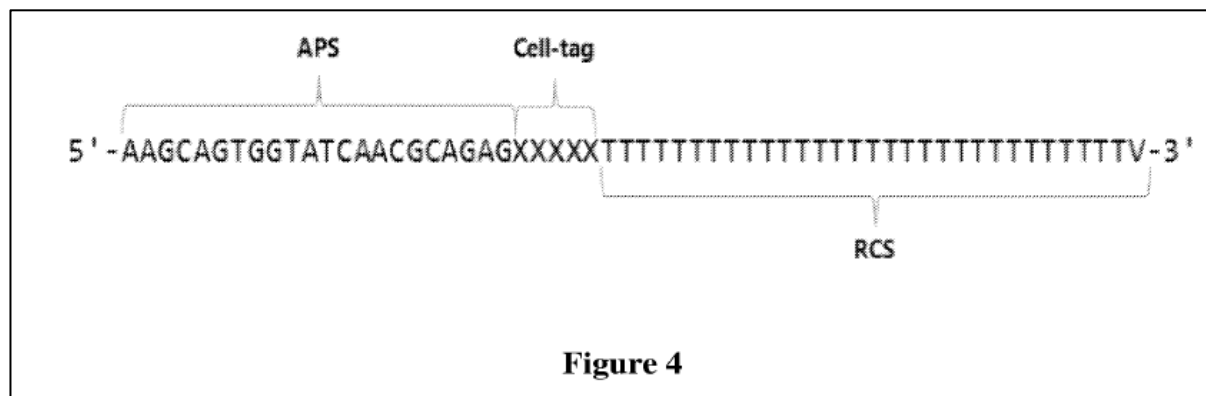
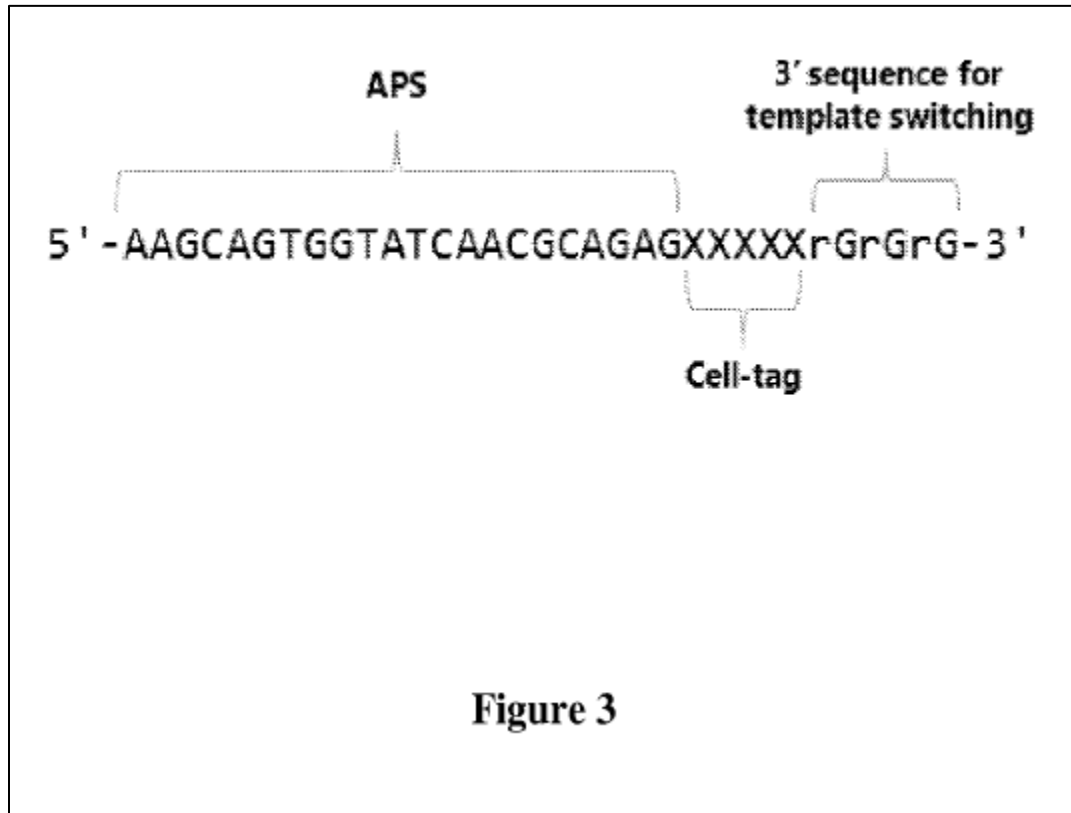
complementary to the TSO.” EX1003, 18:23-24. After the tags are added as part of the reverse transcription process, the cDNA from each well is pooled, and a cDNA library is created, and the “5’ fragment[s]” are sequenced. EX1003, 18:27-19:11.

36. Linnarsson provides two examples of the STRT method—Examples I and II. Example I employs a biotinylated CDS and a TSO with a Bts cut site on the 5’ end of the cell-tag. EX1003, 23:9-19. Figures 2, 3, and 4 depict aspects of the Example I STRT method. EX1003, 23:1-25:4. Figure 2 is a generic overview demonstrating how the CDS anneals to mRNA with a poly-A tail, and TSO anneals to a poly-C sequence of reverse transcribed DNA. Figures 3 and 4 show a “typical” TSO and CDS sequence used in Figure 2. Figure 2 includes no explanation or depiction of purification, sequencing, or any of the other steps persons of skill would recognize as necessary to prepare and subsequently analyze a cDNA library. EX1003, 4:14, 51. Moreover, persons of skill would have understood that the black lines extending from the CDS and TSO in Figure 2 represent STRT method features, including the disclosed TSO and CDS used in Example I, which include the amplification primer sequence (“APS”)³ and Bts cut site for the TSO, and at least the APS, and biotin for the CDS. EX1003, Figs.2-4; EX2015, 1. Further, persons

³ The APS allows the first strand of cDNA to be amplified by PCR. EX1003, 19:18-23.

of skill would have understood that the slanted black line on the CDS indicated a tail, and that the tail included a sequence or molecule that is not complementary to the sequence to which it binds. EX1003, Fig.2. In the context of Example I, persons of skill would have understood this tail in Figure 2 refers to at least the universal primer and the biotin present in Linnarsson's disclosures.





37. Example II describes another example of the STRT method. EX1003, 31:3-32:10. Example II, like Example I, also employs a biotinylated CDS and a Bts cut site on the 5' of the reverse transcribed DNA. EX1003, 32:10-27. Figure 11 is discussed with and depicts steps of a method used in Example II (which are the same

methods used in Example I), and includes details of the process in Panels A-F, which are summarized here. EX1003, 37:31-32, 5:28-6:12;⁴ *see also infra*, ¶¶97-98.

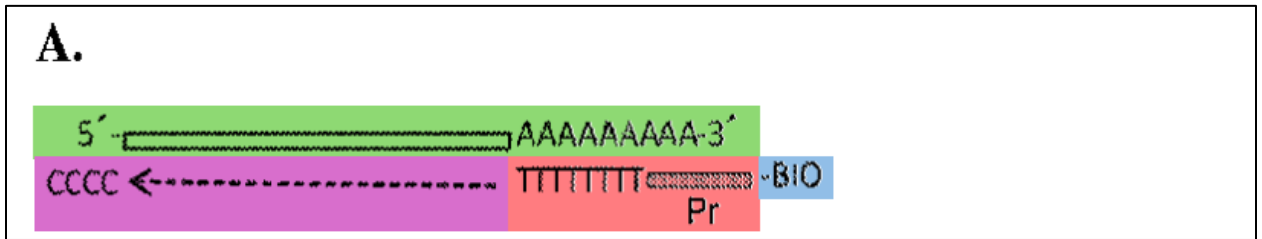
38. **Panel A:** The CDS, in red, has a poly-T tail at its 3' end,⁵ a primer sequence on the 5' side of the CDS's poly-T tail, and is "capped" by a biotin

⁴ The Figure 11 disclosures are also consistent with the STRT method disclosures in Example I. As in Example I, Figure 11 uses a biotinylated CDS and a TSO with a cell-tag and a Bts cut site to generate a 5'-end fragment cDNA library to analyze gene expression of a cell. EX1003, 5:28-6:12, Fig.11. The Figure 11 disclosures are also consistent with Figures 2-4. As described above, Figures 2-4 are simplified depictions of aspects of the CDS and TSO used in Example I (and the same as those used in Example II) and depicted in Figure 11, and would have been understood to include the features of the CDS and TSO disclosed in Examples I and II and Figure 11, including for the CDS the (1) primer (2) Bts cut site (3) cell-tag and (4) biotin and for the TSO the (1) primer (2) Bts cut site and (3) cell-tag. Thus, my opinions on Figure 11 apply equally to Examples I and II and Figures 2-4.

⁵ Each single-stranded nucleic acid has a 5' and 3' end. EX1002, ¶57. These are identified by a phosphate group on the 5' end, and a hydroxyl group on the 3' end.

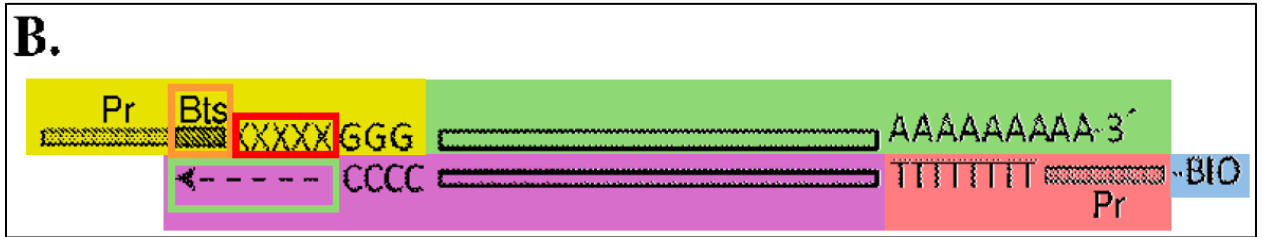
molecule (attached to the 5' deoxyribose phosphate group, preventing the further addition of nucleotides), in blue. EX1003, 5:28-31, Fig.11 (Panel A), 23:13-19; EX2018, 23:21-25:21. A cell-tag (according to Dr. Cooper) can be incorporated between the primer and the poly-T tail. EX1002, ¶107; EX1003, Fig.4, 4:17-18, 18:31-19:11. The CDS's 3' end poly-T tail is hybridized to the mRNA molecule's (green) poly-A tail because Adenine and Thymine are complementary nucleic acids. EX1003, 5:28-31; EX2018, 23:21-25:21. Once the CDS anneals to the mRNA molecule, a reverse transcriptase enzyme recognizes the annealed structure and reverse transcribes the mRNA to create cDNA (purple). See EX1003, 5:28-31, Fig.11 (Panel A). The resulting cDNA (purple) includes an added poly-C tail at its 3' end ("CCCC") due to the terminal transferase activity of reverse transcriptase. EX1003, 5:30-32, Fig.11 (Panel A); EX2018, 25:22-26:6.

Linnarsson refers to the double-stranded ends of the *double*-stranded DNA generated through its process as 5' and 3' "fragments" or "ends." EX1003, 6:5-8, 26:22. The "5' fragment" is the end of the double-stranded DNA corresponding to the region with the TSO sequences, and the "3' fragment" is the end corresponding to the region with the CDS sequences. EX1003, 6:5-8.



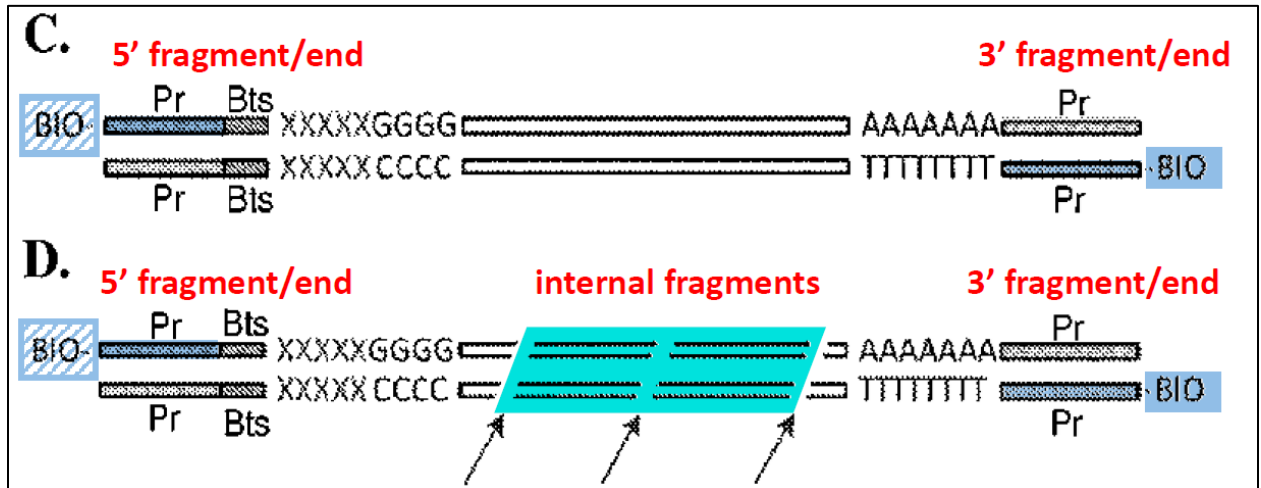
39. **Panel B:** To extend the synthesis of the cDNA (purple) from Panel A, Linnarsson uses a TSO (yellow). EX1003, 5:31-6:1, Fig.11. The TSO's poly-G tail hybridizes to the poly-C tail of the cDNA because Cytosine and Guanine are complementary to each other. EX1003, 5:31-33, Fig.11 (Panel B); EX2018, 25:22-26:12. The TSO includes a cell tag (red box), a Bts sequence (orange box; recognized by the BtsCI type IIS restriction endonuclease), and a primer sequence (Pr) on the 5' side of the poly-G tail. EX1003, 5:31-6:1, 23:13-19, Fig.11 (Panel B); EX2018, 27:14-28:15. A reverse transcriptase enzyme recognizes the annealed structure, and then second-strand cDNA synthesis proceeds using the TSO as a template (depicted by the dotted arrow boxed in green), which adds the cell-tag and Bts to the 3' end of the cDNA (boxed in green) in a second extension step. EX1003, 5:31-33, Fig.11 (Panel B).⁶

⁶ The mRNA molecule is optionally digested by an RNase enzyme. EX1003, 16:10-17. Alternatively, the top strand may be separated from the cDNA during



40. **Panels C and D:** The cDNA sequences generated through the steps shown in Panels A and B are then amplified via PCR, during which a biotin (shaded blue) is added to all cDNA fragments on their 5' ends. EX1003, 25:17-20, Fig.11 (Panel C); EX2018, 28:19-29:7. At this point, the 5' end of each DNA strand is capped by a biotin molecule. EX1003, 25:17-20, Fig.11 (Panel C). Following PCR, the sequences are subjected to a “fragmentation step” with a DNase enzyme, which randomly cleaves the double stranded DNA. EX1003, 6:3-4 (Panel D), 12:4-8; EX2018, 29:8-13. The internal fragments (teal) from Panel D are washed away in a later step (Panel E) (and are therefore never sequenced). EX1003, 6:3-4 (Panel D), 26:26-26.

the PCR process, which includes a step where the double-stranded DNA is separated into single-stranded DNA. EX1002, ¶¶85-86.

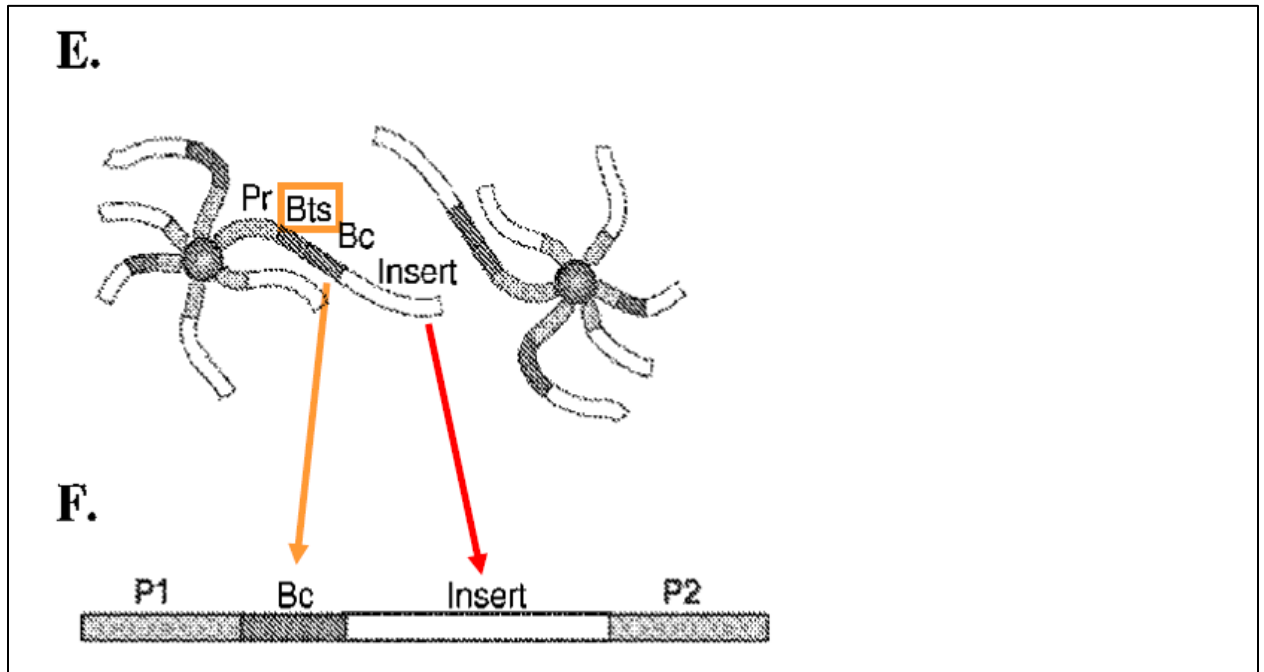


41. **Panels E and F:** The remaining biotinylated and fragmented DNA segments—the 5' fragments with “barcode (Br) and BtsCI site (Bts)” and the 3' fragments (which do not have a Bts site) from Panel D—are then bound to streptavidin beads. EX1003, 6:5-6, 26:21-26, Fig.11 (Panels D-E), 23:13-16, 32:18-27; EX2018, 31:12-22; *see also* EX1003, 21:13-25, 27:6-28:2, 35:23-36:9; EX1002, ¶175; EX2018, 43:21-44:22. It is at this point that the internal fragments from Panel D, which are not biotinylated, and therefore are not bound to beads, are washed away. EX1003, 6:3-4 (Panel D), 26:26-26; EX2018, 43:21-44:22. As I discuss in Paragraph 45, fragmentation may be required in preparation for NGS sequencing. As discussed above, Linnarsson describes no alternative to this step.

42. Following immobilization to the beads for the biotinylated 3' and 5' fragments of the DNA, another enzyme, BtsCI is added, which cleaves the bound fragments at the Bts (orange outline) sequence. EX1003, 6:6-8, Fig.11 (Panel E), 27:6-28:2, 35:23-36:9. Only the portion of the 5' fragments that are to the 3' end of

the Bts sequence are released from the beads; the 3' fragments are biotinylated, and “remain stuck on the beads” because they do not include Bts sequences. EX1003, 6:6-8, Fig.11 (Panel E); *see also* EX1003, 21:13-25, 27:6-28:2, 35:23-36:9; EX1002, ¶175; EX2018, 43:21-44:22. This is confirmed in Panel E, which depicts two different fragments: the first set of fragments has four sequences, and the second set has two sequences. EX1003, Fig.11 (Panel E); *see also* EX1003, EX1003, 21:13-25, 27:6-28:2, 35:23-36:9; EX1002, ¶175; EX2018, 43:21-44:22. The set of fragments with four sequences is, as identified in Panel E, made up of the 5' fragments, and correspond to the side of the DNA tagged with the TSO. EX1003, Fig.11 (Panel E). The set of fragments with two sequences does not contain a Bts sequence and corresponds to the end of the DNA tagged with the CDS. EX1003, Fig.11 (Panel E).

43. Two adapter sequences, P1 and P2, are then ligated to the ends of those 5' fragments released by BtsCI digestion to allow the next-generation sequencer to sequence the fragments (Panel F). EX1003, 6:9-12, 27:6-28:2, 35:28-36:9. If the library is sequenced from the P1 primer, the “reads begin with the 5bp barcode, followed by 3-4 Gs, followed by the cDNA insert.” EX1003, 6:9-12. If the P2 sequenced is used, the “reads would yield only the cDNA sequence.” EX1003, 6:9-12. The 3' fragments are not released, do not have P1 or P2 sequences added, and are not sequenced. EX1003, 6:6-8, 35:28-36:9.



44. The method is depicted below.

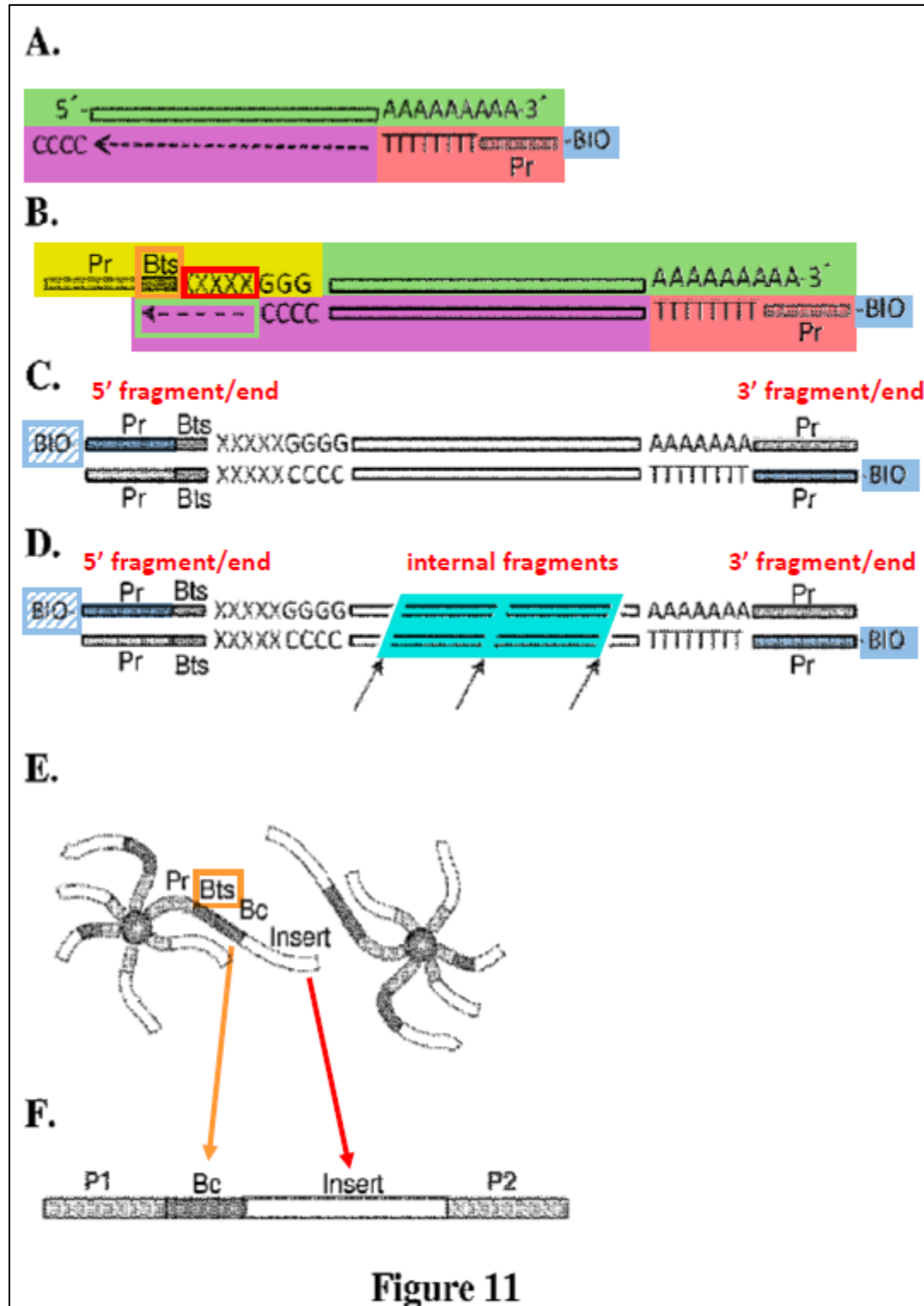


Figure 11

45. Persons of skill would have understood that after the cDNA library is generated, as shown in Examples I and II and Figure 11, further processing steps

(after, *e.g.*, Figure 11 Panel F) are contemplated to prepare it for sequencing, including in NGS. *E.g.*, EX1003, 21:13-25, 48:9-10 (Claim 27). For example, Linnarsson discloses that the cDNA library can be further fragmented, such as for use in shotgun sequencing, which, unlike the process disclosed in Figure 11, would result in each fragment being sequenced (in this case no fragments are “washed away”). EX1003, 22:5-21. Persons of skill would have understood that this further processing would be useful or even necessary depending on the downstream sequencing process being employed. EX1003, 21:13-25, 22:5-21. As another example, further fragmentation might be used to prepare the cDNA fragments such that they are an optimal size for sequencing depending on the sequencing method. EX2020, 57:4-20.

46. Linnarsson’s method results in tagging the mRNA molecules extracted from a single cell with the same barcode through the process of reverse transcribing it into cDNA. EX1003, 3:25-4:2, Fig.11. Thus, after amplification and sequencing of the cDNA, Linnarsson’s method does not allow persons of skill to distinguish, among the multiple copies of polynucleotides, those that are the result of amplification from those that were created endogenously (*e.g.*, multiple copies of mRNA molecules resulting from overexpression of a particular gene), and thus does not determine how many instances of that particular polynucleotide existed *in a particular cell* before amplification. *See* EX1003, 22:22-28. Linnarsson’s method

also does not sequence the entirety of the DNA sequences. Not only are the internal fragments washed away, but the 3' ends are not released from the beads and are not sequenced. EX1003, 6:5-8.

47. As is clear from the consistent, repeated disclosures of Linnarsson, the features disclosed in Examples I (including Figures 2-4) and II (including Figure 11)—Linnarsson's only provided examples—both affirmatively use CDS and/or TSO, biotin, a restriction site (*e.g.*, the Bts cut site), and fragmentation, which are essential features of Linnarsson's methods. EX1003, Figs. 2-4, 11, Examples I-II; EX2015, 1. Figure 4, for example, depicts a polynucleotide sequence with an APS, a cell-tag, a poly-dT RCS, and a Bts cut site. EX1003, Fig.4; EX2015, 1. Persons of skill would understand Figure 4 as illustrating additional sequence details of a typical CDS used in Figure 2, and not as attempting to illustrate other molecules (like biotin), which the text confirms are part of the Example I process. The CDS is designed to target the mRNA molecules in the cell, and triggers reverse transcription to generate the strand of cDNA with the poly-C tail. EX1003, 5:30-31, 15:1-11, 25:2-22, Figs.2, 11 (Panel A). The TSO then binds and triggers the generation of the rest of the strand of cDNA. EX1003, 16:18-17:2, 18:19-26, Figs.2, 11 (Panel B). Together, these two reverse transcription steps ensure the tags and PCR primers are added. EX1003, Figs.2, 11 (Panel C). And the biotin ensures the tagged fragments are bound to the beads, and not washed away with the internal fragments

and untagged polynucleotides the researcher is not interested in sequencing (e.g., DNA in Linnarsson's method). EX1003, 21:13-25, 35:23-36:9. This is how Linnarsson discloses purifying its sequences that will become the cDNA library. The Bts cut site located on the 5' fragments is then recognized by the Bts enzyme, and when cut releases the 5' fragments for clean-up and sequencing following purification. See EX1003, 12:4-8 (defining "cDNA library" as "DNA (cDNA) fragments."). This is further confirmed by the absence of any alternative approaches in Linnarsson for performing the STRT method.

2. McCloskey (EX1004)

48. EX1004 is an article by Megan L. McCloskey, Reinhard Stöger, R. Scott Hansen, and Charles D. Laird, titled "Encoding PCR Products with Batch-stamps and Barcodes." EX1004 ("McCloskey"), 1. McCloskey is focused on addressing laboratory molecular contamination issues for irreplaceable samples. EX1004, 1.

49. As part of the DNA amplification process, McCloskey adds an "encoding oligonucleotide" with a "batch-stamp" and barcode, to copies of alleles for a particular targeted gene. EX1004, 2-4. McCloskey's encoding oligonucleotide includes, *inter alia*, (1) a batch-stamp; and (2) a random barcode that distinguishes alleles of the same gene with (3) a genomic target-binding site that targets the

specific gene on the chromosomal DNA to be barcoded prior to amplification. EX1004, 2. The batch-stamp identifies the batch sample from which the polynucleotides originated in order to identify contamination (*i.e.*, to identify sequences originating from a different patient or sample left over from a prior analysis) and the barcode distinguishes between copies of an allele for a particular gene in the sample (without identifying the particular cell of origin). EX1004, 2-4.⁷ The only specific length taught in McCloskey for its random barcodes is seven nucleotides, and that barcode is comprised of the four natural deoxyribonucleotides.⁸ EX1004, 4. This allows for the creation of up to 16,384 (4^7) unique barcodes and provides enough tag sequences to tag different alleles for the same gene differently in McCloskey's collection of cells. EX1004, 4.⁹

⁷ Alleles are different versions of the same gene. EX2003, 3; EX1002, ¶60.

⁸ "Deoxyribonucleotides" are nucleotides that have a deoxyribose as their sugar component. EX1002, ¶¶55-56; EX2014, 31; EX1001, 7:23-48. The four natural deoxyribonucleotides are adenine (A), guanine (G), cytosine (C), and thymine (T). EX1002, ¶¶55-56; EX2014, 31.

⁹ I agree with Dr. Cooper that there are, at most, two alleles per "normal" cell. EX1002, ¶60.

50. McCloskey's method involves lysing the cells and combining the contents (including the DNA) of those cells before tagging the copies of a particular gene within this combined material to determine the presence of alleles. EX1004, 2, 5. Among other things, this is confirmed by the amount of DNA used in the described experiment. EX1004, 5. The 2.71µg used is significantly more genomic DNA than that present in a single cell. EX2016 ("6.46 pg" mean weight in a diploid cell); EX2018, 34:19-35:3. 2.71µg of DNA corresponds to approximately 420,000 cells'-worth of DNA ($2.71 \mu\text{g} \div 6.46 \text{ pg} = 419,504.64$). EX2016. In my opinion, this amount of DNA in McCloskey indicates many cells were lysed, the genetic material combined, and then isolated.

51. After this material is combined, copies of a particular gene (including any alleles or mutations) are tagged by annealing the genomic target-binding site of the oligonucleotide to the complementary DNA sequence found on one end of the target gene. EX1004, 5. McCloskey discloses that the presence of an encoding oligonucleotide in the DNA for the particular gene of interest allows researchers to identify and remove contaminating sequences (from different batch samples) and redundant sequences (amplified copies containing the same random barcode) after the gene of interest is amplified. EX1004, 3 (Fig. 1), 7. However, McCloskey's method does not allow a user to determine whether or which copy of a particular allele (or mutation) is present in a given cell because *no copies of the target allele*

are tagged with a sequence that tracks an allele to its cell of origin. EX1004, 3. Single-cell analysis requires separate tagging of polynucleotides for each cell *prior* to pooling to ensure that the same source identifying tag is introduced only to nucleic acids from the same cell. EX1002, ¶71 (“(1) isolation of single cells from a cell population; (2) template preparation and tagging of sequences...”). But McCloskey pools the nucleic acids from all of the cells in the sample *prior to tagging*, making it impossible to track a nucleic acid to its cell of origin. EX1004, 2, 5. Thus, unlike the Challenged Claims, McCloskey’s approach to removing redundant copies simply permits a user to determine whether the target gene was present, including whether there were any alleles or mutations present, *in the sample when it was tagged*. EX1004, 2, 5.

52. McCloskey’s disclosures are focused on tagging specific genes, the sequence of which is already known; they are not focused on tagging all mRNA. In my opinion, there would not be enough oligonucleotide tags in McCloskey to tag each mRNA molecule in a cell. EX1003, 1; EX1004, 4. As discussed above, persons of skill would have expected there to be between 50,000 and 300,000 mRNA molecules in the typical mammalian cell, and persons of skill would want to err towards having enough unique tags for the high side of that range. EX1054, Abstract. McCloskey’s barcode, with just 16,348 unique combinations, is far too few to allow unique barcode tags for each mRNA molecule. There would not have

been enough unique tags for even the 50,000 mRNA molecules (the low end of the range of expected mRNA molecules) to be tagged, even under ideal conditions (where each unique tag is present and anneals to a mRNA molecule), let alone the many multiples of tags that would have been required to tag those 50,000 molecules. *Supra*, ¶28.

3. McCloskey II (EX1055)

53. EX1055 is U.S. Patent Publication 2007/0020640, titled “Molecular Encoding Of Nucleic Acid Templates for PCR And Other Forms Of Sequence Analysis.” EX1055 (“McCloskey II”).

54. McCloskey II focuses on addressing laboratory molecular contamination issues for irreplaceable samples by tagging specific genes with known sequences. EX1055, ¶¶3-5, ¶9.

VI. Claim Construction

55. I understand that, for purposes of *inter partes* review, claim terms are construed in accordance with their ordinary and customary meaning. I understand claim construction begins with the claim language itself. Claim terms are presumed to have their ordinary and customary meaning in light of the patent’s specification as understood by persons of skill at the time of invention, unless (1) the patentee sets out a definition and acts as his own lexicographer, or (2) the patentee disavows the

full scope of a claim term either in the specification or during prosecution. I understand that the “ordinary and customary” meaning of a claim term is the meaning that the term would have to persons of skill at the time of the invention, and that in addition to the claim language, the terms are also understood in the context of the patent’s specification and file history.

56. I have assumed the following construction in my analysis, which is found in the specification of the ’197 Patent: “sample” means “a quantity of material from a biological, environmental, medical, or patient source in which detection, measurement, or labeling of target nucleic acids is sought” (EX1001 (’197 Patent), 12:7-10).

57. For all other terms, I have been asked to assume that they have their plain and ordinary meaning as they would have been understood by persons of skill in light of the specification.

VII. THE CHALLENGED CLAIMS ARE VALID

A. Ground 1: Linnarsson In View of McCloskey Does Not Render Claims 1-4, 6-11, 20 or 23-25 Obvious.

58. In reviewing Dr. Cooper’s declaration, I found his argued combination of prior art disclosures unclear, including as it relates to Element 1F. EX1002, ¶¶157-171. I reviewed the Rehearing Decision in connection with this Element and understand the Board cited Dr. Cooper’s combination of conceptually dividing the

nucleotides of Linnarsson's cell-tag into two parts, with one part (six to eight of the nucleotides) being used for the functionality of McCloskey's "batch-stamp," which Dr. Cooper asserts is the "first tag sequence," and the second part (seven of the nucleotides), being used for the functionality of McCloskey's barcode, relying on Dr. Cooper's Paragraph 168. Rehearing Decision, 13-14, 17. I will interpret Dr. Cooper's theory consistent with the Board's for purposes of my opinions below. (To the extent Dr. Cooper or Petitioner may be permitted in the future to argue some different interpretation, I reserve the right to respond further.)

1. Dr. Cooper Fails To Show Linnarsson Teaches "Sequencing The Tagged Polynucleotide"

59. I understand that Parse, relying on Dr. Cooper for support, has argued in its Petition that "sequencing is performed on the entire tagged polynucleotide, which includes the sample sequence and both tag sequences." Dr. Cooper asserts that the "tagged polynucleotide" in Linnarsson is the cDNA generated during reverse transcription, and that, in his opinion, these tagged polynucleotides are sequenced. EX1002, ¶136 ("The tagged cDNA is the claimed 'plurality of tagged polynucleotides.'"), ¶¶172-175. In my opinion, Dr. Cooper's analysis does not establish that Linnarsson teaches sequencing "the entire tagged polynucleotide" as Parse argues for limitation [1G]. Dr. Cooper ignores, and therefore does not address, that Linnarsson discloses sequencing of only *a portion* of each tagged

polynucleotide, as opposed to the entirety of the tagged polynucleotides as is required by Parse's argument. *Supra*, ¶34.

60. Linnarsson teaches that the “amplified library is fragmented to 200 – 300 bp using controlled DNase digestion,” which creates 5', 3', and internal fragments. EX1003, 6:3-6, 21:13-25, 27:6-28:2, 35:23-36:9; EX1002, ¶175; EX2018, 43:21-44:22; *supra*, §V.B.1. The internal fragments are portions of the sample polynucleotide. EX1003, Fig.11. Both ends of the “tagged” cDNA molecules are biotinylated (at the 5' end of each strand of cDNA), and therefore both end-fragments of double-stranded DNA “are bound to beads” after fragmentation. EX1003, 6:3-6; *supra*, § V.B.1. Linnarsson discloses that the “***internal fragments are washed away***” before sequencing. EX1003, 6:3-6; *id.* at Figs.11 (Panels C-E); *supra*, § V.B.1. As explained above, this is because the internal fragments are not biotinylated, and, as a result, are not bound to streptavidin beads. *Supra*, §V.B.1. Thus, because they are washed away, these portions of the sample polynucleotides are not available to be sequenced.

61. After the internal fragments are washed away, the 5' and 3' fragments remain. The 5' fragments correspond to the portions of the cDNA molecules with the TSO sequences, and the 3' fragments correspond to the portions of the cDNA molecules with the CDS sequences. *Supra*, §V.B.1. Linnarsson states that an enzyme, BtsCI, is added, which releases the portion of the 5' fragments on the 3'

end (the end further from the bead) of the Bts cut site from the beads to which these fragments were bound. EX1003, Fig.11 (Panel C-E). As explained above, the 3' fragments do not contain a Bts sequence and, as a result, *are not released from the “beads”* to which they are bound. EX1003, 6:3-6; *id.* at Figs.11 (Panels C-E); EX2018, 43:21-44:22; *supra*, §V.B.1. Thus, Linnarsson does not disclose sequencing the 3' fragments of the “tagged polynucleotide,” which remain bound to the beads. EX2018, 43:21-44:22. The only portion of any “tagged polynucleotide” that Linnarsson discloses sequencing is the 5' fragment portion. *Supra*, §V.B.1.

62. My paragraphs 60 and 61, immediately above, reflect how a person of skill would understand Linnarsson's STRT method—the method Dr. Cooper relies on for preparing a cDNA library for sequencing. EX1002, ¶¶105, 108, 112, 137, 138, 141, 150-151, 196, 199, 200 (same); EX2018, 23:15-25:21, 29:14-30:2; EX1003, 5:10-11, 23:1-4, 37:31-32, Fig.11. Linnarsson's STRT method discloses reverse transcribing mRNA to prepare a cDNA library, which can then be sequenced using NGS, and Linnarsson does not disclose sequencing without fragmentation (or fragmenting the tagged polynucleotide and then sequencing each resulting fragment). EX1003, 6:3-12 (“fragmentation step”), 12:4-8 (“**The term ‘cDNA library’ refers to a collection of cloned complementary DNA (cDNA) fragments...**”), 21:13-22 (“the cDNA library can be processed by fragmenting the cDNA library (e.g., with DNase) to obtain a short-fragment 5'end library”), 23:15-

20, 26:9-20, 35:16-36:9, Fig.11. Thus, it is my opinion that Dr. Cooper failed to show that Linnarsson discloses “sequencing” the “entire tagged polynucleotide.”

63. In Dr. Cooper's deposition, he asserted that Figure 11 is described as “optional steps that can be used within STRT” and that “fragmentation is one optional step” in that process. EX2018, 23:4-14, 30:3-12. I acknowledge the description of Figure 11 as a whole says “optional,” but disagree that the *fragmentation* shown in Figure 11 is optional, including because Linnarsson discloses no other way of preparing the cDNA library for sequencing, nor does Dr. Cooper identify any. *Supra*, ¶47. And, to the extent Dr. Cooper is arguing that he showed in his report that Linnarsson teaches sequencing of the entirety of each tagged polynucleotide, I disagree with that as well. First, persons of skill would understand that Linnarsson's consistent disclosure of fragmenting then sequencing indicates that only fragments are sequenced in Linnarsson. (Linnarsson's repeated reference to sequencing the 5' fragment, without any disclosure of sequencing any other fragment, further confirms this understanding.) Second, Dr. Cooper's three citations regarding sequencing do not support an argument that Linnarsson teaches sequencing the full length of tagged polynucleotides. EX1002, ¶¶174-175. Linnarsson at 4:8-13 is a description of Figure 1, which is a high-level overview of

Linnarson's STRT method.¹⁰ EX1003, 4:8-13. That introductory description does not indicate what steps are taken to sequence *any* cDNA, let alone whether all or some of the tagged polynucleotides are sequenced. EX1003, 4:8-13. Persons of skill would not make assumptions about what is sequenced without reading the *rest* of Linnarsson's description of its STRT method, which, in context, indicates that the sequencing is of just the 5' fragment. Further, Linnarsson at 19:28-32, also cited by Dr. Cooper, explains that the "5' APS can be designed to facilitate downstream processing of the cDNA library" and that it "can be designed to be identical to the primers used in [certain] sequencing methods." EX1003, 19:28-32. But this statement, too, does not indicate that the entire tagged polynucleotide is sequenced. Dr. Cooper's final citation in connection with his assertion that Linnarsson discloses sequencing the tagged polynucleotide directly supports my opinion that the entirety of the tagged polynucleotide is *not* sequenced in Linnarsson. EX1002, ¶175 (Citing EX1003, 21:13-25. The cited portion of Linnarsson states that the "cDNA library can be processed *by fragmenting the cDNA library* (e.g. with DNase) *to obtain a short-fragment 5'-end library.*" EX1002, ¶175 (citing EX1003, 21:13-25). This citation explains that the cDNA library (created after fragmentation, as shown in,

¹⁰ It also refers to "the tissue of interest." which persons of skill would have understood refers to a eukaryotic multi-cellular organism.

e.g., Figure 11) may be further processed (including by additional fragmentation) to make it suitable for high-throughput sequencing. This disclosure similarly fails to indicate the “entire tagged polynucleotide” is sequenced. Indeed, Linnarsson does not disclose any other approach to processing the cDNA library for sequencing, and its only descriptions include fragmentation. EX1003, Fig.11, 26:9-20 (Example I’s “DNase treatment”), 35:17-27 (Example II “was fragmented by DNase 1”); EX2018, 21:21-22:12.

64. Relatedly, both Dr. Cooper’s CDS theory (in which both tag sequences are included in the “conceptually divided” cell-tag located on the CDS (EX1002, ¶168)), and Dr. Cooper’s “alternative[]” theory (in which one of the two tags is included in the CDS (EX1002, ¶170)) fail for the additional reason that the CDS is part of the 3’ fragment, and Dr. Cooper identifies no teaching in Linnarsson showing the 3’ fragment is sequenced. EX1003, Fig.11 (Panel E); EX1003, 6:8; EX2018, 43:21-44:22. Although the 3’ fragment (the end of the cDNA with the CDS) is not part of the internal portions that are washed away, it is nevertheless not sequenced. *See* EX1003, Fig.11 (Panel E); EX1003, 6:8 (“3’ fragments remain stuck on the beads.”). And nowhere does Linnarsson describe releasing any 3’ fragment from the bead or sequencing it. EX2018, 43:21-44:22. And under Dr. Cooper’s “alternativ[e]” theory, Dr. Cooper identifies no way to associate the 3’ fragment with the corresponding 5’ fragment even if both were released. EX1002, ¶¶136, 170. In

my opinion, it is not clear how this would be accomplished based on the approach of Linnarsson. Thus, Dr. Cooper's theories relying on the CDS of the Linnarsson-McCloskey combination fails because any tags in the CDS would be on the 3' fragment, and would not be sequenced.

65. Thus, Dr. Cooper's Ground 1 fails because Dr. Cooper failed to show that Linnarsson (including Linnarsson as modified by McCloskey) teaches sequencing the "entire tagged polynucleotide," which is how I understand Parse argued this limitation is met.

2. McCloskey Does Not Disclose A "First Tag" Sequence

66. Dr. Cooper asserts in his "conceptually divided" theory that McCloskey's batch-stamp functions as a cell tag (first tag sequence), and that modifying Linnarsson's cell tag with McCloskey's batch-stamp renders claim 1 obvious. EX1002, ¶168. I disagree. In my opinion, this modification does not render claim 1 obvious because McCloskey's batch-stamp is not a "first tag sequence" as required by claim element [1E], [1G] and [1H].

67. The claimed "first tag sequence" identifies the cell from which a nucleic acid originated. EX1001, Cl.1. McCloskey's batch-stamp does not perform this function. Rather, McCloskey teaches that its batch-stamp allows one to distinguish between nucleic acids arising from all the cells *in the sample*, on one hand, and the contaminant nucleic acids arising from cells *from different, prior samples*, on the

other hand. EX1004, 3-4. Confirming this, McCloskey further teaches that each *sample* uses the same batch-stamp, regardless of the sample size, so nucleotides either have the batch stamp (if they are from *any* of the cells in the sample) or have a different distinct batch-stamp if they are contaminants from a previously analyzed sample. EX1004, 4 (“We typically use between six and eight base pairs to encode sample identify and date of reaction.... The assigned batch-stamp is registered for a specific sample run on a specific date and is used only once.”); EX2018, 35:4-19. This is the process by which McCloskey identifies contamination. EX1004, 4.

68. In asserting that McCloskey’s batch-stamp can somehow be a cell-tag, Dr. Cooper points to the use of the phrase “single cell” (specifically, “single cell... analyses” and “single cell sample”) in McCloskey to attempt to suggest McCloskey teaches “single cell analysis” and read a cell-tag (which must be different for each different cell in a sample) into McCloskey. *See* EX1002, ¶159. In my opinion, at most, McCloskey discloses the possibility of tagging DNA even when the sample (batch) consists of DNA from only a “single cell,” not actual “single cell analysis” (which involves analyzing multiple cells from the same sample while maintaining information on a cell-by-cell basis). EX1004, Abstract, 7. Single cell analysis is the study of polynucleotides and proteins from multiple cells at a cell-by-cell resolution, which preserves the unique information of each cell in the sample, such as the protein content of each cell. *See* EX1002, ¶68. In the ’197 Patent, such analysis relies on

cell tags so that it is possible to distinguish amongst polynucleotides from different cells after combining them together for, *e.g.*, sequencing. McCloskey does not disclose using its method for “single cell analysis”—rather, at most it discloses it for use on a one-cell sample. EX1004, Abstract, 7; *supra*, ¶64. Tagging a “single-cell sample” would not require uniquely tagging polynucleotides from different cells to identify the cell from which they originated as there is only one cell in the sample to be tagged. Tagging different cells in the same sample differently would be inconsistent with McCloskey’s use of a batch-stamp anyway, which uses the same batch-stamp for the entire sample. I address the two phrases involving “single cell” in McCloskey further below.

69. McCloskey’s reference to a “single-cell sample” (EX1004, 2) is not “single cell analysis,” nor is it within the scope of the claims, which require the sample consist of “a **plurality** of cells.” EX1001, Cl.1. Indeed, persons of skill would understand McCloskey to be referring, literally, to a sample of a single (*i.e.*, one) cell. Further, a “first tag sequence” (cell tag) would not be used to distinguish among cells in a sample when the entire sample is only a single cell. However, McCloskey’s batch-stamp is useful for “single-cell sample[s]” as contamination is of particular concern when the initial sample is so small and significant amplification is required. EX1004, 1. Amplification bias for a single-cell sample can be significant because many rounds of PCR are required to have enough material to

sequence when starting with only the DNA for two alleles from a single cell. Further, McCloskey provides additional context confirming what it means with respect to “single-cell samples” in discussing the authors’ prior work in Miner on a “single-cell sample”: “[The Miner] approach is highly successful in identifying valid, double-stranded sequences and their methylation patterns, but requires bisulfite conversion and *more template material than is usually available from irreplaceable or single-cell samples.*” EX1004, 2. Persons of skill would have understood this was because the risk of contamination and amplification bias are of particular concern when the entirety of the sample consists of one cell. This is consistent with McCloskey’s citations to Miner (EX1013), Laird (EX2007), and Li (EX2019). I have reviewed each of these references and none discloses “single cell analysis.”

70. McCloskey’s references to “single cell... analyses” (as ellipsized by Dr. Cooper) also would not be understood by persons of skill to refer to “single cell analyses,” nor does it somehow show that McCloskey teaches tagging material from different cells differently. EX1002, ¶160. One of the cited McCloskey excerpts reads: “We recommend that batch-stamps and barcodes be used when amplifying *irreplaceable* DNAs and cDNAs available for forensic, clinical, single cell, and ancient DNA analysis.” EX1004, Abstract, 7. A “single cell” is an irreplaceable sample, but “single-cell analysis” is not. *Id.* “Single cell...analysis” in McCloskey

would be understood by persons of skill to relate to the disclosure of “single-cell samples” discussed above, and would be understood as relating to the analysis of a sample of a single cell.

71. That McCloskey would have been understood *not* to disclose “single cell analysis” is further confirmed by the fact that McCloskey does not disclose a working method that could be used to perform a single cell analysis. McCloskey includes no sequence that distinguishes among cells from the same sample, nor does it disclose isolating cells such that each tagged polynucleotide would be able to be traced back to its cell of origin. McCloskey’s encoding oligonucleotide’s batch-stamp allows one, instead, to distinguish between nucleic acids arising from a cell *in the sample* and contaminant nucleic acids arising from cells *from prior samples*. EX1004, 3-4. Similarly, McCloskey’s encoding oligonucleotide’s barcode allows one to distinguish *between nucleic acids arising from the same sample*, but provides no information as to which allele sequences originate from the same cell, or whether the different sequences originate from different cells. EX1004, 4. Indeed, McCloskey pools all the DNA from all the cells in a sample *prior to tagging*, which would make it impossible to determine which cell a particular allele arose from and is incompatible with so-called single cell analysis. EX1004, 2, 5; EX2018, 33:3-35:3. Persons of skill would not have understood McCloskey’s passing reference to “single cell . . . analyses” to refer to anything other than samples consisting of a

“single cell,” and Dr. Cooper’s misreading of McCloskey fails to support a reason to combine the references.

3. McCloskey’s Batch-Stamp Lacks The Benefits Of Linnarsson’s Cell-Tag For Single-Cell Analysis

72. I disagree with Dr. Cooper’s opinion in connection with his “conceptually divided” theory that persons of skill would have been motivated to modify Linnarsson’s cell-tag to incorporate the functionality of McCloskey’s batch-stamp and barcode. EX1002, ¶168. In my opinion, persons of skill would not have been motivated to modify Linnarsson’s cell-tag to incorporate the functionality of McCloskey’s batch-stamp because Linnarsson’s cell-tag was specifically designed for use in single-cell analyses, and includes a further beneficial feature designed to identify and eliminate the misattribution of data from a first cell to a second caused by transcription errors in the cell-tag.

73. Linnarsson’s method for single-cell analysis includes a specifically designed cell-tag adapted for single-cell analysis. Linnarsson discloses that “[o]nly exact barcodes were allowed, and the barcodes were designed so that no single error would convert one valid barcode into another.” EX1003, 36:28-29. In my opinion, this is an important feature of Linnarsson that persons of skill would have appreciated. Linnarsson’s barcodes identify the cell of origin. EX1003, 32:23-34 (Table 9). A transcription error during amplification or reverse transcription that

changed the barcode sequence would impact the analysis in one of two ways. Either (1) there would be a sequence that belonged to none of the 96 cells tagged with a barcode or (2) the error would cause a sequence to be misattributed to a different cell of origin. This first error is easily identifiable, as any sequences with a barcode matching none of the 96 barcodes employed in the experiment could be eliminated. However, the second error, where an error causes a first barcode sequence to become one of the other 95 barcodes employed, would impact the gene expression analysis by distorting whether and to what extent any cell is expressing a particular gene. Linnarsson's cell-tag barcodes are intentionally designed to prevent this error from impacting the results. EX1003, 36:28-29.

74. However, McCloskey's batch-stamp teachings, aimed at an entirely different purpose, include no such feature. EX1004, 4. Instead, McCloskey aims to identify and eliminate contaminants from outside the sample altogether, and as a result does not distinguish among the cells from which material within the sample originated to begin with, and also is not concerned with the impact these types of transcription errors may have.

75. McCloskey's batch-stamp is intended to eliminate contaminant sequences after the nucleic acids (alleles) are tagged with oligonucleotides containing the same batch-stamp. EX1004, 1-7. Thus, after the sample alleles are tagged and amplified, any transcription error within the batch-stamp is of no

moment—any batch-stamp that is “not valid” (*i.e.*, does not match the batch-stamp used in that particular sample)—is simply eliminated from the dataset. EX1004, 3 (Fig.1.). There would be numerous other allele copies amplified from the original template nucleic acid that, after sequencing and elimination of the redundant copies or copies with transcription errors, would provide the information McCloskey’s method is designed for—*i.e.*, whether a particular allele copy is present in the sample. EX1004, 1. McCloskey was not concerned with, nor did it contemplate, the scenario where a transcription error may cause the sequence to be misidentified as being part of a different sample. EX1004, 1-7.

76. Dr. Cooper never acknowledges this beneficial aspect of Linnarsson’s already-existing design, nor does he offer a modification to McCloskey’s batch-stamp teachings that would maintain this feature, which persons of skill would have recognized as particularly useful in Linnarsson’s method. EX1002, ¶¶157-171. For these reasons, it is my opinion that persons of skill would not have been motivated to modify Linnarsson’s source-identifying tag with a tag employing McCloskey’s batch-stamp approach, thus eliminating the benefits of Linnarsson’s cell-tag.

4. There Are Not Enough Unique Tags In McCloskey To Tag “Each Polynucleotide” In Linnarsson

77. Dr. Cooper’s primary theory for adding a second tag to Linnarsson is based on his assertion that “a POSA could have conceptually designated the first 6-

8 nucleotides of Linnarsson’s cell-tag sequence to identify the single-cell origin of the tagged polynucleotide, and use the next 7 nucleotides to distinguish the tagged polynucleotide from other one [sic] from the same single cell.” EX1002, ¶168; *see also supra* ¶58. Dr. Cooper’s “alternativ[e]” theory also uses the seven-nucleotide barcode from McCloskey, but instead of including it in the CDS that contains the cell-tag, this theory includes the barcode in the TSO (or if the cell tag is in the TSO, the barcode would be in the CDS). EX1002, ¶170 (“this alternative design would be identical to the method taught by Linnarsson except that either the TSO or CDS would be modified such that it is different for each tagged cDNA”); ¶168 (“a POSA could have conceptually designated the first 6-8 nucleotides of Linnarsson’s cell-tag sequence to identify the single-cell origin of the tagged polynucleotide, and use the next 7 nucleotides to distinguish the tagged polynucleotide from other one from the same single cell.”). Thus, both theories rely on McCloskey’s seven-nucleotide barcode. Dr. Cooper asserts for both of his theories that there is a motivation to combine Linnarsson with McCloskey because McCloskey “provides a more definitive solution” and “solve[s]” Linnarsson’s alleged problem of amplification bias. EX1002, ¶166. According to Dr. Cooper, this amplification bias is resolved in his theories “[b]y incorporating a second tag in an oligonucleotide that uniquely identifies *each sample polynucleotide* prior to amplification.” EX1002, ¶166 (“By incorporating a second tag in an oligonucleotide that uniquely identifies each sample

polynucleotide prior to amplification, McCloskey's method is able to uniquely account for each individual sample polynucleotide and 'solve' amplification bias"), ¶91 (*"To solve this problem [of amplification bias]... Every sample and the polynucleotides within every given sample are first uniquely tagged before being amplified..."*). EX1002, ¶91.

78. Dr. Cooper also asserts that "a POSA would have understood that [Linnarsson's] method was designed to be applied to substantially all mRNA molecules of a given cell." EX1002, ¶222.¹¹ I agree with Dr. Cooper. Persons of skill would have understood that the purpose of the disclosures of Linnarsson on which Dr. Cooper relies, including the design of Linnarsson's tags, and the purpose of Linnarsson's method, is to tag substantially all mRNA molecules of a given cell. For example, Dr. Cooper relies on Linnarsson's disclosures of a CDS with a poly-dT sequence, which would indiscriminately tag the eukaryotic mRNA present in the cell. See EX1002, ¶¶104, 105, 107, 108, 110, 112, 136-139, 141, 150, 151, 154, 158, 196, 199-200, 222.

¹¹ While this opinion is included in Dr. Cooper's claim 25 analysis, claim 25 depends from claim 1, and his interpretation of Linnarsson is relevant to the analysis of both claims.

79. For these reasons, I disagree that Dr. Cooper's proposed combination would be able to solve this alleged amplification bias problem. In my opinion, persons of skill would not have thought McCloskey's 7-nucleotide-long barcode would provide enough unique combinations to tag "each sample polynucleotide"—all mRNA of a mammalian cell—in Linnarsson's method to eliminate amplification bias as Dr. Cooper argues. EX1002, ¶¶167-168. Thus, persons of skill would not have been motivated to make the modification proposed by Dr. Cooper.

80. In my opinion, McCloskey does not disclose enough unique tags to tag "each sample polynucleotide."¹² McCloskey's disclosed barcode is "seven nucleotides" long and uses one of the four natural deoxyribonucleotides in each

¹² Dr. Cooper opined that "Linnarsson's method then tags these sample polynucleotides for the purpose of gene expression analysis. *A POSA would understand that the purpose of such a method would be to analyze substantially all mRNA molecules of a cell—therefore, this method would provide associating a tag with substantially every sample mRNA molecule. . . . Thus, a POSA would understand that Linnarsson describes a method which is designed to be applied to substantially all mRNA molecules of a given cell.*" EX1002, ¶222. Therefore, I interpret the "sample polynucleotides" of claim 1 to include "substantially all mRNA molecules of a given cell."

position. EX1004, 4; *see also* EX1002, ¶168. McCloskey teaches that each of the seven positions in its barcode can be composed of one of four deoxyribonucleotides, resulting in 16,384 unique seven nucleotide-long barcodes.¹³ EX1004, 4 ($4^7 = 16,384$). For Linnarsson's purposes, a barcode of that length is insufficient to distinguish among each of the vast number of mRNA molecules present in a mammalian cell. "In general, to accomplish such molecular tagging the number of distinct MID tags to be used should be many times greater than the actual number of molecules to be analyzed." EX1001, 22:54-60. Dr. Cooper asserted here that one would want "*at least 100 times [more] than the number of sample polynucleotides.*" EX1002, ¶253. For purposes of my opinion, I have assumed this is true (but this does not alter my opinion: even if fewer unique tags were needed per polynucleotide, McCloskey's barcode would still not provide enough). *See also*, ¶79.

81. In contrast to McCloskey's analysis of alleles, Linnarsson is focused on analyzing gene expression by measuring and identifying the mRNA present in a cell.

¹³ These barcodes are typically generated randomly. *E.g.*, EX1035, 1, 3-4. Persons of skill would understand that this is a theoretical maximum, unlikely to be achieved in practice, and therefore the actual number of unique tags in McCloskey is less than 16,348.

EX1004, Abstract. Persons of skill would expect anywhere from 50,000 – 300,000 mRNA molecules to be present in the typical mammalian cell addressed by Dr. Cooper. EX1054, Abstract; EX1002, ¶255. McCloskey's barcodes, which are intended to distinguish between a small set of *alleles* (one to two per cell), would be insufficient to distinguish among each of the vast number of *mRNA* molecules present in such a mammalian (or other eukaryotic) cell. EX1054, Abstract; EX1002, ¶255. If Dr. Cooper's assertion that 100 unique tags would be needed for every sample polynucleotide to be tagged is assumed, for example, then at least 30,000,000 unique tags would be required to have enough to tag the expected mRNA. EX1002, ¶¶253, 255. Persons of skill would not have known how many mRNA molecules would be present in a cell before the sample is analyzed, and would therefore have used enough unique tags to assure that each mRNA molecule could be tagged even if the number of mRNA molecules in the cell were at the high end of the expected range.¹⁴ EX1002, ¶256; EX1054, Abstract.

82. It is also my opinion that tagging the mRNA molecules in Linnarsson's gene expression analysis method using a seven nucleotide barcode as taught in

¹⁴ Based on Dr. Cooper's assumptions, even for the low end of the range, persons of skill would need at least 5,000,000 unique tags, which is far greater than the 16,348 McCloskey's tag may generate.

McCloskey would not resolve amplification bias. At most, only a subset of the mRNA molecules could be tagged using the no more than 16,348 barcodes taught by McCloskey. Dr. Cooper argues that resolving amplification bias requires tagging “[e]very sample and the polynucleotides within every sample.” EX1002, ¶91. But tagging only a subset of the mRNA molecules with unique tags would not be tagging “each” or “every” polynucleotide. EX1002, ¶¶91, 166. It also would not allow persons of skill to determine the number of molecules in the sample because persons of skill would not have been able to determine which of the *untagged* molecules were original to the sample or were a result of amplification bias.

83. Thus, Dr. Cooper has not shown that persons of skill would have been motivated to modify Linnarsson’s method for analyzing *gene expression* to include a barcode where that barcode is deficient for that purpose and there would not be enough barcodes to “uniquely identif[y] each sample polynucleotide” as Dr. Cooper claims is the motivation for Petitioner’s proposed combination. EX1002, ¶166.

5. McCloskey Does Not Describe Single-Cell Analysis

84. Dr. Cooper asserts that McCloskey and Linnarsson both “describe methods of genetic analysis of single cells” and are “thus in the same field of endeavor” in arguing it would be obvious to combine the two references for both of

his theories. EX1002, ¶163. I disagree that McCloskey discloses single-cell analysis for the reasons identified above.

85. Further, in my opinion, even if McCloskey's use of the phrase "single cell" or "single-cell samples" did refer to "single cell analysis," persons of skill would not have been motivated to use McCloskey's teachings in "single cell" **gene expression** analyses, such as Linnarsson's. McCloskey's approach was designed for tagging particular *alleles* in a sample, of which there are only, at most, two per cell. EX1004, 4. What Dr. Cooper argues is McCloskey's general references to using its batch-stamp and barcode analysis for "single cell" analysis does not mean McCloskey was suitable for "single cell" *gene expression* analysis, or any other of the many types of single cell analyses one may do. EX1004, 4; EX1003, 1. Persons of skill would have understood that McCloskey's design was not suitable for analyses of very large numbers of nucleic acids such as Linnarsson's (*e.g.*, 50,000-300,000 mRNA molecules). EX1002, ¶222. In addition, McCloskey lacks a tag that identifies the cell of origin, which is also necessary for single-cell analysis. For these reasons, persons of skill would not have been motivated to use McCloskey's teachings for the type of analyses in Linnarsson.

6. Dr. Cooper's Claim 1 Reasonable Expectation of Success Fails

86. In my opinion, persons of skill also would not have had a reasonable expectation of success in “tagging each polynucleotide” with Dr. Cooper’s proposed combination of Linnarsson and McCloskey under either of his theories. EX1002, ¶¶166, 168, 222. As explained above, persons of skill would not have had reasonably expected to be able to tag each sample polynucleotide with a different barcode by incorporating McCloskey’s barcode teachings into Linnarsson because McCloskey would not have provided enough unique barcodes to do so. *Supra*, §VII.A.4. As discussed above, McCloskey’s 16,384 tags are an insufficient number of unique tag sequences to tag the 50,000 – 300,000 mRNA molecules in a cell even under an ideal scenario where each tag hybridizes to a sample polynucleotide. And it is far less than the 5,000,000 – 30,000,000 unique tags that would be required, based on Dr. Cooper’s assumptions, to tag each of those mRNA molecules in a real-world application. *Supra*, §VII.A.4.

B. Ground 2: Linnarsson In View Of McCloskey In View Of Knowledge Of A POSA Does Not Render Claim 5 Obvious

1. Dr. Cooper's Theory

87. Dr. Cooper’s Ground 2 theory does not indicate (1) whether or not the tags in the CDS or TSO from Ground 1 remain in the CDS or TSO in Ground 2; (2) whether the ligated oligonucleotide tag includes just the second tag sequence or the

first and second tag sequence; (3) at what step in Linnarsson's method the ligation occurs; and (4) how the ligation is achieved. Given these uncertainties, I have done my best to respond to Dr. Cooper's potential theories below.

2. Dr. Cooper's Motivation To Modify Dr. Cooper's Ground 1 Fails

88. As discussed above, it is unclear what tags are argued to be added via ligation in Dr. Cooper's Ground 2 theory. To the extent Dr. Cooper's theory involves modifying Ground 1 by *adding one (or two)* tags via ligation to the CDS and/or TSO, without removing the tags already included in the CDS and/or TSO, it is my opinion that person of skill would have had no motivation to make this modification. *See supra* §VII.B.1 (discussing lack of clarity in combination).

89. Dr. Cooper's motivation to combine fails. For example, if I assume Dr. Cooper's theory involves adding tag(s) to the CDS or TSO and that the CDS or TSO from Dr. Cooper's Ground 1 theory remain in the CDS or TSO, then in my opinion persons of skill would not have been motivated to add more tags (and more steps) when the CDS or TSO already contains two tags. Dr. Cooper provides no supported motivation for adding redundant tags, and persons of skill would not increase the complexity, cost, or time required to practice the method by adding redundant tags.

90. If, on the other hand, under Dr. Cooper's theory, the tag(s) are removed from Dr. Cooper's Ground 1 theory, then Dr. Cooper identifies no teaching in any

reference showing how to remove tag(s) from these nucleic acids, nor does he identify any reason persons of skill would have removed the tags to instead add them back by ligation.

91. Regardless of whether Dr. Cooper's theory involves ligating tag(s) that allegedly *replace*(s) something in the CDS or TSO or *adds* tag(s), ligating a tag *adds steps* to the Ground 1 method and alters its principle of operation. Linnarsson is designed to generate a sequencing library of tagged cDNA via annealing CDSs and TSOs, one or both with the alleged cell tags, which are carried forward through the transcription process. *Supra*, §V.B.1. In my opinion, persons of skill would not have been motivated to fundamentally alter that process by, *e.g.*, removing the functionality of those tags from the CDS or TSO and then adding tags to the fragments at some point after extension—not only because this would fail to achieve the intended purpose (*infra*, §§VII.B.3.b, VII.B.4)—but also because it is a fundamentally different approach that rejects the teachings of Linnarsson that Dr. Cooper uses as his starting point.

92. It is also my opinion that persons of skill also would not have been motivated to change and complicate the Linnarsson method (as modified in Ground 1) only to yield what Dr. Cooper asserts is an “equivalent” process. Dr. Cooper's Ground 2 theory would require persons of skill to, at minimum, buy or create another set of tags as well as any other materials necessary to implement the ligation (such

as Hug's linker molecules and enzymes). This would also require adding at least one additional ligation step at some point in the process, and, to the extent Dr. Cooper's theory involves the ligated tags replacing any tags that are in the CDS or TSO, removing those allegedly existing tags from Linnarsson's CDS or TSO. In my opinion, persons of skill would not have added steps simply to ligate tags rather than use existing tags in the TSO or CDS, as Dr. Cooper asserts persons of skill would do in Ground 1. Persons of skill would not add tags and then remove them without using them so that different tags could be added later. Certainly, persons of skill would not have added such steps merely to achieve an "entirely equivalent" method. EX1002, ¶230.

93. Regardless of Dr. Cooper's precise combination, Dr. Cooper attempts to justify modifying Linnarsson by asserting "it would allow for increased flexibility in selecting tag sequences during an experiment." EX1002, ¶230. First, Dr. Cooper's assertion about ligation providing "extra flexibility" is contradicted by Dr. Cooper's assertion about the method being "entirely equivalent." EX1002, ¶230. Second, Dr. Cooper's assertion is entirely conclusory, and, in my opinion, is wrong. EX1002 ¶230. Dr. Cooper has not identified any "flexibility" that is gained when *ligating* a tag rather than, *e.g.*, adding it via primer extension, which Dr. Cooper asserts would have been obvious *before* further modifying its combination with ligation. Assuming one could add a tag via ligation in Linnarsson's method (*infra*,

§VII.B.3), those same tag sequences could be added via primer extension with the CDS and/or TSO.

94. The lack of benefit is confirmed by Dr. Cooper's theories. The same tag(s) that Dr. Cooper proposes adding via ligation would serve the same purpose as the ones allegedly included in the CDS or TSO in Ground 1. Dr. Cooper does not identify any tags that could be added by ligation that would not have been included in the CDS or TSO under his own Ground 1 theory. EX1002, ¶230.

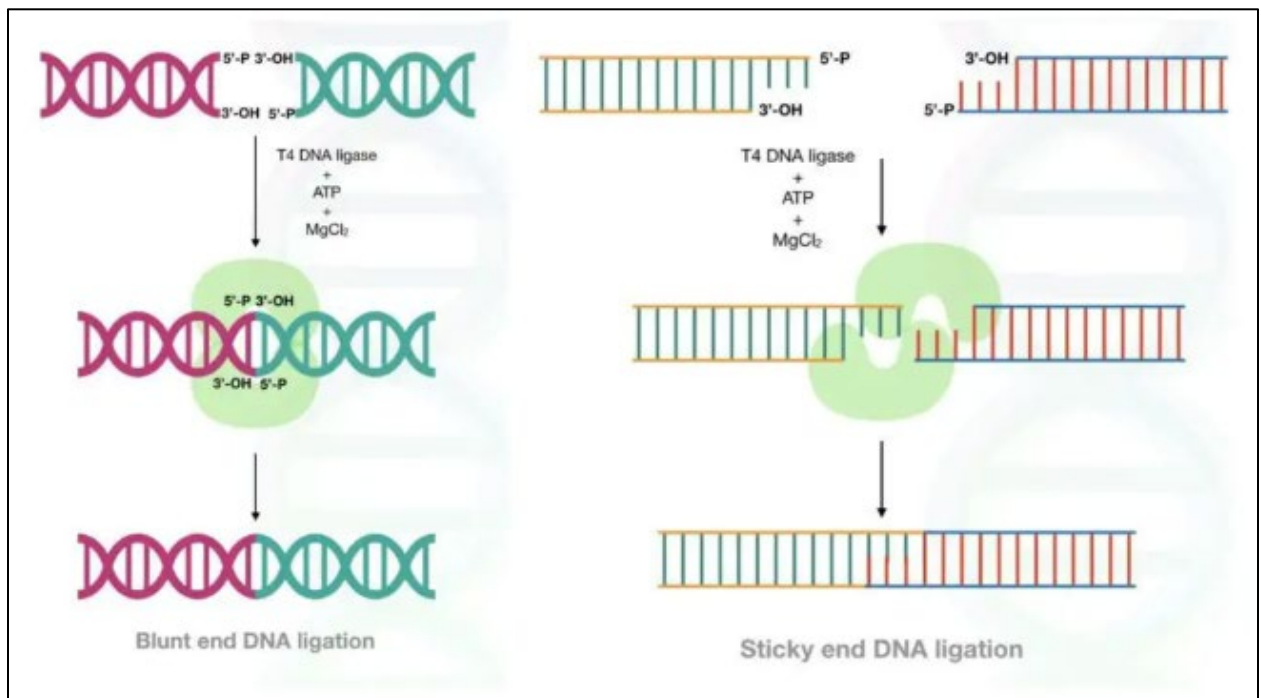
3. Dr. Cooper Has Not Shown A Tag Can Be Successfully Added Via Ligation In Linnarsson While Still Practicing The Claimed Method

a) Dr. Cooper Did Not Show That A Tag Can Be Ligated To The CDS

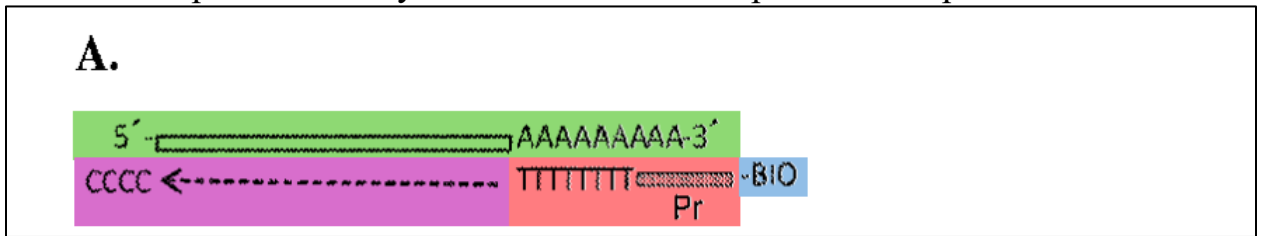
95. In my opinion, as the combination is presented by Dr. Cooper,¹⁵ *ligating* a tag to the CDS is incompatible with the other steps in Linnarsson's method, and would not result in *a method that practices claim 5*. EX2020, 41:22-44:20 (Dr. Cooper testifying that his Ground 2 theory only involves ligating one tag).

¹⁵ Dr. Cooper does not mention, let alone explain, any other modifications to its combination that would make it possible to ligate any tag to the CDS.

96. First, Linnarsson teaches that a biotin is attached to the 5' end of the CDS. *Supra*, §V.B.1. The biotin on the 5' end of the CDS would block ligation of a tag to the CDS. EX1003, Fig.11, 5:28-6:8, 23:13-14. Ligation “means to form a covalent bond or linkage between *the termini of two or more nucleic acids*.” EX1001, 6:15-32. This occurs when the ends of DNA fragments are joined by the formation of phosphodiester bonds between the 3'-hydroxyl of one DNA terminus with the 5'-phosphoryl of another. EX1001, 6:23-40. Both of the two forms of ligation, depicted in the picture below, result in formation of phosphodiester bonds between adjacent nucleotides. The attached biotin would prevent ligation because the biotin would be attached to the 5' phosphate and thus would block the formation of a phosphodiester bond required for ligation.

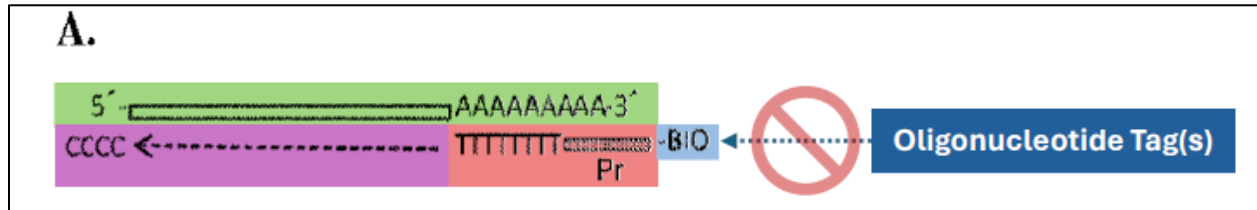


EX2017. An example of a biotinylated nucleic acid is depicted in the picture below.



EX1003, Fig.11. Linnarsson's biotin molecule is bound to the 5'-phosphoryl group in the cDNA strand. EX1003, Fig.11, 23:13-14, 25:17-18, 32:18-25. This is both depicted in Figure 11 and is disclosed in the CDSs used in Examples I and II, which persons of skill would understand to mean that the biotin is bound to the 5' end of the nucleic acid, which has a phosphoryl group on the end of the last nucleotide. *Id.* This means that a nucleic acid tag cannot be covalently bound to the 5' phosphoryl group as that bond is occupied by the biotin. *Id.* As shown in Figure 11 and described in Examples I and II, the CDS is capped by a "BIO" (biotin) molecule at its 5' end. EX1003, Fig. 11; 23:13-15 (The sequence of STRT-**BIO** (which is a CDS) is: 5'-**BIO**-AAGCAGTGGTATCAACGCAGAGT₃₀VN-3'"); 32:19-20 ("STRT-T30-**BIO** (5'-**biotin**-AAGCAGTGGTATCAACGCAGAGT₃₀VN-3'")), 32:18-27 (disclosing the same CDSs and TSOs for Example II); EX2018, 23:15-25:21, 29:14-30:2. Thus, the tag cannot be ligated to the "CDS" because the biotin blocks the necessary covalent bond from being formed with its 5' end. For this reason, it is my opinion that Dr. Cooper's CDS theory fails. Dr. Cooper offers no evidence that ligation could occur to a molecule with a 5' biotin, and one is left only with his

unsupported (and incorrect) assertion, which ignores the existence of the biotin in Linnarsson's approach, as the only "evidence" that one would have any expectation of success. Given the basic physical configuration of the molecules, it is hard to imagine how one might carry out Cooper's proposed reaction.



97. I disagree with Dr. Cooper's deposition testimony implying that the use of biotin is optional in Linnarsson's method. EX2018, 25:10-21, 31:1-22; EX2020, 10:20-11:2; *supra*, §VII.A.1. Dr. Cooper expressly relies on Figure 11 (disclosed in the context of Example II), which persons of skill would have understood requires biotin. EX1002, ¶¶105, 108, 112, 137, 138, 141, 150, 151, 196, 199, 200; *supra*, ¶62. Although Linnarsson describes other aspects of its method as optional, Linnarsson never describes biotin as optional, and the only examples of Linnarsson's method (Examples I and II) include the use of biotinylated CDSs and biotinylating the "5' end" during PCR. *See* EX1003, Fig.11, Example I, Example II. And Dr. Cooper (whose report never suggests eliminating biotin in any case) does not address how Linnarsson's method would work *without* biotinylated cDNA. For example, Linnarsson discloses that the biotin captures the tagged fragments and enables their purification. EX1003, 5:28-6:12, Fig.11. The fragments *without* biotin, described

as the “internal fragments,” are “washed away.” Without biotin, all the fragments would be “washed away.” *Id.* Moreover, in Linnarsson the biotin allows these sequences to be attached to the streptavidin beads, which allows these sequences to be isolated such that adapter sequences may be added, and then sequenced. EX1003, Fig.11.

98. Second, these CDS sequences used in Examples I-II and Figure 11 ***do not include a Bts*** sequence. EX1003, Examples I-II, Fig.11 (Panel A). The Bts sequence is key to Linnarsson’s method, as it is the only disclosed approach for releasing the 5’ fragments of cDNA from the streptavidin beads and proceed to the sequencing step. EX1003, 5:28-6:12, 27:6-28:2, 35:18-36:9, Fig.11; *Infra*, ¶102. Therefore, these CDS sequences, corresponding to the 3’ fragments, would remain bound to the beads in Panel E and would not be sequenced since only sequences with the Bts sequence are released by BtsCI digestion, and Dr. Cooper identifies no teaching in Linnarsson of releasing the CDS sequence from the beads. EX1003, 6:4-8 (“3’ fragments remain stuck on the beads.”), 27:6-8, Fig.11; EX2018, 43:21-44:22. Thus, any tag allegedly ligated to the CDS would not practice the method of dependent claim 5 because these sequences would not be sequenced, as required by independent claim 1. EX1001, Cl.1.

99. And, finally, even if the tag were ligated to the biotin (such that it were cDNA-biotin-tag), and some portion were somehow released from the beads, it is

unclear how the tag would be released with the 3' fragment. The biotin group would somehow be *between* the cDNA sequence and the added tag, and thus at least one sequence (the second tag or the 3' fragment) would not be released by Bts digestion. Further, if both fragments were somehow released (if, for example, a Bts sequence was added to both sides of the biotin), the fragments would no longer be connected, and the second tag could not be used to identify the 3' fragment, as required by claim 1. EX1001, Cl.1.

100. For all these reasons, it is my opinion that Dr. Cooper's theory of ligating a tag to the CDS fails.

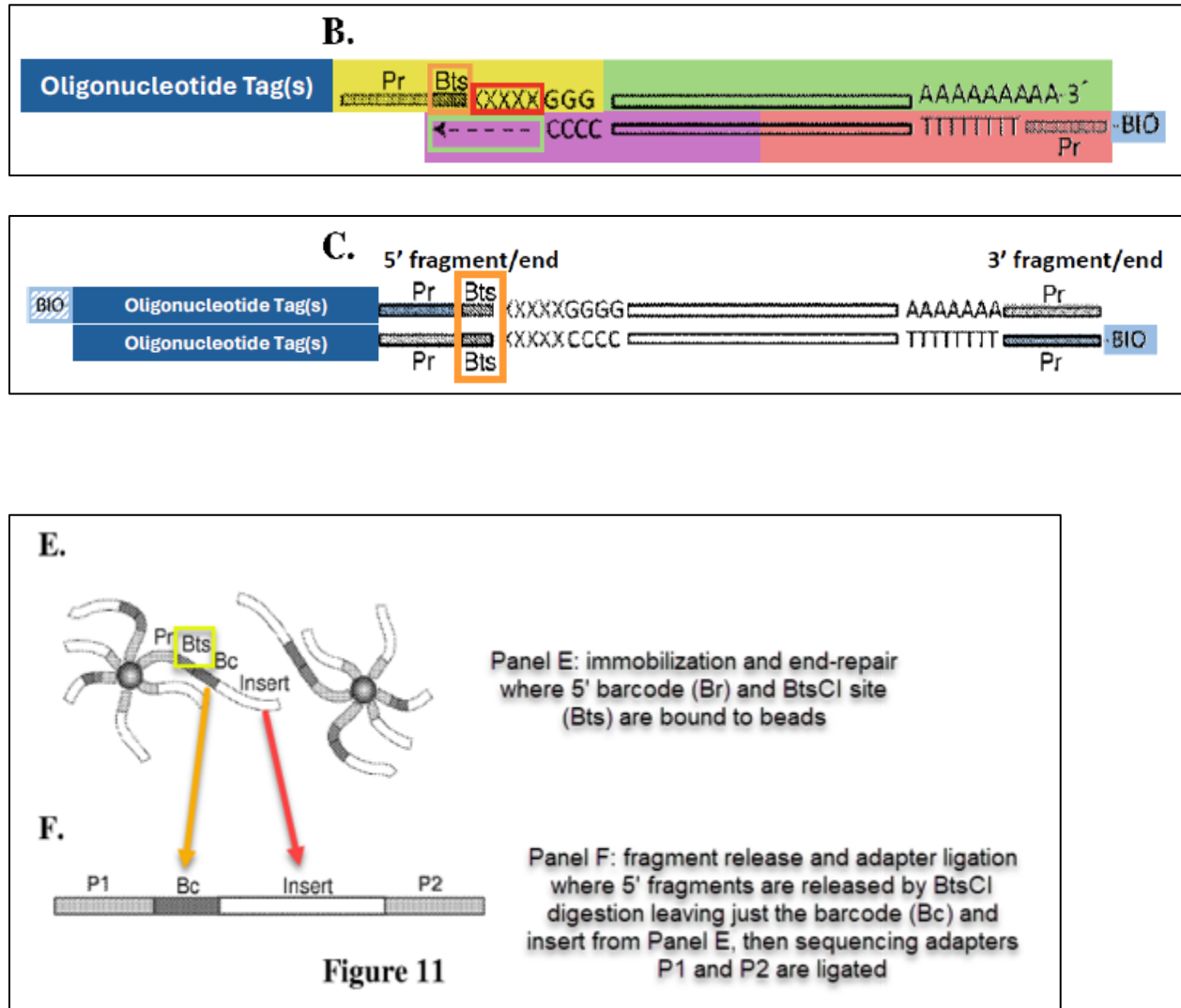
b) Linnarsson Teaches That Any Tag Ligated To The TSO Will Be Cleaved Off Prior To Sequencing

101. It is additionally my opinion that the tag also cannot be ligated to the TSO to arrive at a method that practices claim 5 under Dr. Cooper's theory. Dr. Cooper asserts the tag would be ligated to the end of the TSO "after extension." EX1002, ¶230. Dr. Cooper's theory fails for at least two reasons. First, ligating the "second tag" to the TSO "after extension" would mean that the ligation occurs after the template switching step in Panel B (the next extension step following the addition of the TSO) is completed. EX1003, Fig.11. At this point in Linnarsson's method, the cDNA has been extended twice, and is complete. There is no third "extension" step to reverse transcribe this newly ligated sequence to the resulting cDNA.

EX1003, Fig.3. In Panel A, the cDNA is extended a first time to generate complementary sequence to the mRNA. EX1003, Fig.11 (Panel A). In Panel B, the cDNA is extended a second time to include a sequence complementary to the TSO. EX1003, Fig.11 (Panel B). There is no disclosed “third extension” of the cDNA in Linnarsson that would result in any sequence complementary to the ligated tag if the tag were ligated “after extension.” Thus, this theory fails because the tag added via ligation would not be included in the cDNA that is subsequently amplified during PCR in Panel C, and therefore would not be part of the tagged polynucleotide that is sequenced in Panel F. EX1003, Fig.11.

102. Second, even if the tag were somehow reverse transcribed, Dr. Cooper’s TSO theory fails for another reason. The ligated tag would be located on the 5’ end of the Bts sequence, shown in the annotated Panels B and C below. EX1003, Fig.3, Fig.11; *see also id.* 27:6-7, 35:28-36:9. But according to Linnarsson, any sequence on the 5’ side of the Bts sequence will be **removed**—*i.e.*, “released by BtsCI digestion, **leaving just the barcode (Bc) and insert**” that are released from the streptavidin bead for sequencing. EX1003, 5:28-6:12, Fig.11 (Panel F); *see also id.* 27:6-7, 35:28-36:9. Thus, even if the tag were ligated to the end of Linnarsson’s TSO, Dr. Cooper ignores that the tag would be **cleaved off** from the sample polynucleotide in Panel E. EX1003, 5:28-6:12, Fig.11. Once cleaved, there would be no “second tag sequence” (as required by claim element [1F]) on the released

polynucleotide, and thus the “tagged polynucleotide” would not be sequenced (as required by claim element [1G], nor would it be possible to distinguish among sample polynucleotides as required by claim element [1I], from which claim 5 depends. EX1001, cl.1.



4. Dr. Cooper’s Other Ligation Assertions Fail

103. In my opinion, neither Linnarsson nor McCloskey teaches ligation of a “second tag sequence” to generate the tagged polynucleotides. Dr. Cooper also does

not assert that ligation of a tag to a polynucleotide is taught by either reference. EX1002, ¶¶222-231.

104. Dr. Cooper's Ground 2 is "Linnarsson in View of McCloskey, and further in View of the Knowledge of a POSA." EX1002, §XII.B. However I note that Dr. Cooper relies on EX1007, titled "Measurement of the Number of Molecules of a Single mRNA species in a Complex mRNA Preparation, by Hubert Hug and Rainer Schuler" ("Hug"), when analyzing Ground 2 and arguing the obviousness of the ligation limitation of claim 5. Even to the extent Hug is considered in analyzing Ground 2, I disagree that claim 5 is obvious.

105. First, it is my opinion that persons of skill would not have looked to Hug's fundamentally different (and theoretical) approach or been motivated to use its linker approach with Linnarsson or McCloskey, let alone in a Linnarsson-McCloskey combination. For example, Hug does not sequence the polynucleotides to which it attaches linkers. Hug uses linkers that hybridize to each other to randomly add a number of identical linker units to a molecule that had been cleaved by a BamH1 enzyme so that, rather than DNA-sequence based barcodes, Hug's barcodes are based on sequence-length (or number of repeated linker elements). EX1007 (Hug), 7-8, Fig.3 ("addition of linkers with a defined ration of oligonucleotide terminators to create DMS of different lengths"). These linkers are then ligated to each other through the use of an enzyme, T4 DNA Ligase. EX1007

(Hug), 7-8, Fig.3 (“ligation with T4 DNA ligase”). Hug then analyzes the DNA fragments by length in order to distinguish between individual molecules and to then estimate the number of molecules in a sample.¹⁶ EX1007 (Hug), 7-8, Fig.3 (“determination by PAGE”) (“DNA length analysis by PAGE”); EX2018, 49:8-50:9. This is a fundamentally different approach to analyzing nucleic acids from those used in Linnarsson or McCloskey, which both rely on sequencing the nucleic acids they have tagged. EX1003 (Linnarsson), 6:9-12, Fig.11; EX1004 (McCloskey), 6-7. It is also an approach that persons of skill would have been motivated to avoid using in a method where the nucleic acids were sequenced because it would add a significant number of nucleic acid bases as linker-based bar codes, which would then need to be sequenced. This would be a significant disadvantage when using sequencing methods in which only a small number of base pairs can be sequenced—each “wasted” base added by Hug would represent one fewer base that could be read in the mRNA sequences, and thus possibly mean one misses meaningful sequence variants or fails to distinguish transcript isoforms. The short sequence reads characteristic of “Next Generation Sequencing” (NGS) in use

¹⁶ Hug does not have a MID. These linkers are not “unique,” cannot be used to identify source, and are not sequenced so as to permit differentiation among nucleic acids. EX1001, 6:33-7:22.

at the time are why Linnarsson teaches fragmenting the tagged polynucleotides with DNase to “200 – 300 bp.” EX1003, 6:4-6, 22:5-21. Longer sequences would not be sequenced by many NGS technologies, and longer fragment lengths would therefore be disadvantageous, as persons of skill would have recognized.

106. Second, it is my opinion that Hug’s linker molecule could not be used with the random barcode of McCloskey. Dr. Cooper incorrectly asserts that Hug’s “linker” would have been “*hybridize[d] with the TSO and/or CDS and tag sequences to align them such that they could be ligated together using routine ligation techniques.*” EX1002, ¶229. This is incorrect because Hug’s disclosed linkers are not designed to link a series of random nucleotides, like those in McCloskey’s barcode, with another sequence, like Linnarsson’s CDS or TSO, and Dr. Cooper offers no modifications to any of the molecules to allow such linking.

107. Hug describes adding a BamH1 recognition site to the generated double-stranded DNA such that when it is cut with BamH1 (a particular endonuclease), the resulting double-stranded DNA has a 5’ sticky end with a particular overhanging sequence. EX1007, 7-8, Fig.3. Two different sets of short linkers are subsequently added. EX1007, 7-8, Fig.3. The first set of linkers consists of the recognition site of the BamH1 restriction enzyme and an end with a short

polynucleotide sequence that hybridizes to the other linker.¹⁷ EX1007, 7-8, Fig.3. These linkers can form a chain of a random length because they can continuously hybridize to each other. *Id.* The second set of linkers are “Terminator Sequences,” which terminate the chain because no linkers have the complementary sequence necessary to hybridize to the terminator sequence to continue the chain of linkers. EX1007, 7-8, Fig.3. This is the process by which Hug proposes to differentiate between the sample mRNA molecules. EX1007, 7-8, Fig.3. Hug’s theory is that, using the randomly generated linker chains of various *lengths*, one can differentiate between mRNA molecules originating in the sample and estimate the number of each that were endogenous to the sample. EX1007, 7-8, Fig.3. And Hug’s theory is premised on the first set of linkers being able to hybridize to each sample polynucleotide that was cut with BamH1. EX1007, 7-8, Fig.3. There is no “selectivity” in the linkers, and the method requires that the same BamH1 site be present. EX1007, 7-8, Fig.3.

108. Dr. Cooper ignores these details of Hug’s ligation method and ignores the fact that Hug’s linkers are not designed to accommodate oligonucleotide barcodes with *unique sequences*, such as McCloskey’s barcodes, or designed to

¹⁷ Unlike tags that distinguish polynucleotides (EX1001, cl.1), Hug’s linkers are not unique.

hybridize to Linnarsson's CDSs or TSOs, neither of which contain a BamHI recognition site. EX1007, 7-8, Fig.3; EX1003, Fig.11; EX1004, 2-4. For a linker to work with McCloskey's barcode, there would have to be a linker molecule that is complementary to each combination of sequences that could be present in McCloskey's barcode, and there must be enough of each to ensure the linkers hybridize with the barcode.

109. Moreover, the other end of the linker would need to have a sequence complementary to Linnarsson's CDS or TSO such that it actually "links" the TSO and/or CDS with McCloskey's barcode. In other words, there must be a linker for each of the possible barcode sequences. And for each of these unique linkers, there must be enough of each to ensure that the linker "finds" the barcode and hybridizes to it, and then hybridizes to a CDS and/or TSO. Again, none of this was addressed by Dr. Cooper.

110. And, there would be further issues with this approach because if a linker binds to the CDS or TSO, and another linker binds to the barcode, those two linkers would not then be able to hybridize to each other, and therefore there would be potentially many sample polynucleotides that are never tagged, or insufficiently tagged (*e.g.*, only tagged with a linker, which is unable to act as the second tag). Dr. Cooper does not explain how this would be achieved, particularly in view of the

issues below, or why persons of skill would have been motivated to make this modification, which introduces numerous issues, as outlined above.

111. Alternatively, if Hug's linkers were not modified, McCloskey's barcodes and Linnarsson's CDS and/or TSO would have to be modified to include a BamH1 recognition sequence. Dr. Cooper does not propose this modification, or explain how it would be achieved or why persons of skill would have been motivated to make this modification.

112. Further, even if persons of skill were to have looked at Hug and made whatever modifications were necessary to allow Hug's linkers to work with McCloskey's barcodes and Linnarsson's CDSs and/or TSOs, it is my opinion that Hug does not cure the other technical issues with Ground 2. Hug, for example, does not address ligation to the end of a polynucleotide that is capped with a biotin. Regardless of whether any modifications to the Linnarsson's CDS, Hug's linker, or McCloskey's barcode were made that would permit Hug's linker to hybridize to both, the CDS includes a biotin at its 5' end. *Supra*, §VII.B.3.a. As explained above, no molecule could be ligated at the end of a biotinylated nucleic acid. *Supra*, §VII.B.3. Similarly, Hug does not change where the Bts sequence is located on the TSO or add one to the CDS. *See supra*, §§V.B.1, VII.B.3.a-b. Thus, even if Hug's linkers ligated McCloskey's barcode to the CDS or TSO, (1) the barcode would not be released from the bead for sequencing if ligated to the CDS because there is no

Bts sequence, or, (2) if ligated to the TSO, the barcode would still be cleaved off from the portion of the polynucleotide that was sequenced because it is on the 5' end of the Bts sequence. *Supra*, §VII.B.3.a-b.

5. Dr. Cooper's Ligation Theory Fails If Ligation Occurs After Amplification

113. As stated above, Dr. Cooper's report never specifies *when* the ligation he posits is supposed to occur. Dr. Cooper admitted this in his deposition on the '981 Patent, which included the verbatim argument Dr. Cooper makes regarding ligation here. EX2018, 54:8-21. In my opinion, Dr. Cooper has failed to show obviousness for this additional reason, because his stated theory permits the tag to be ligated after amplification. If the sequences are amplified and then tagged, each amplified sequence (most of which were not endogenous nucleic acids) would have a unique "second tag" sequence, preventing the user from determining which sequences were a result of amplification, and therefore would prevent the user from determining how many of each sequence were endogenous to the sample. In my opinion, this would not meet limitation (d) of claim 1: "using the first tag sequence and the second tag sequence to count a number of sample polynucleotides in said plurality of sample polynucleotides of said cell." EX1001, Cl.1. Each would appear to be an endogenous polynucleotide, making it impossible to correlate the "identified

polynucleotide sequence with the sample polynucleotide from which the identified polynucleotide sequence is derived.”

114. I also disagree with Dr. Cooper’s assertion that Paragraph 231 of his declaration “explicitly” states that the second tag would be ligated prior to amplification. EX2020, 37:11-39:11. Paragraph 231 states: “Thus, it is my opinion that it would have been an obvious design choice to a POSA to ligate the second tag sequence to the TSO and/or CDS to generate the tagged polynucleotide.” In my opinion, PCR generates polynucleotides, and if the polynucleotides which are to be amplified are tagged, then PCR would also generated tagged polynucleotides. Therefore, Dr. Cooper’s assertion that Paragraph 231 explicitly states the timing requirement is incorrect, as it does not explicitly require that the ligation occur prior to amplification.

C. Ground 3: Linnarsson In View Of McCloskey In View Of McCloskey II Does Not Render Claim 13 Obvious

115. Dr. Cooper does not address claim 1 in his Ground 3 theory, and it is thus unclear how the elements of claim 1 are satisfied in the dependent claims of Ground 3. I have been asked to assume that Dr. Cooper’s Ground 3 arguments build on his Ground 1 claim 1 analysis. For the reasons I stated in Ground 1, I disagree that Ground 3 renders any claim obvious.

116. Moreover, I note that the motivation Dr. Cooper identifies in his Ground 3 analysis—“to ensure that each sample polynucleotide is tagged with a unique tag sequence”—is the same motivation Dr. Cooper articulated in Ground 1—to “remove amplification bias by introducing barcodes to measure true sample complexity/template diversity. By tagging each sample polynucleotide uniquely, true sample complexity can be analyzed because each unique sample sequence can be identified by the tag and amplification bias can be accounted for and corrected.” EX1002, ¶167. However, Dr. Cooper does not explain why Ground 1 would be further modified or improved.

VIII. CONCLUSION

117. In summary, for the reasons described above, I have concluded that each of the Challenged Claims is not invalid as obvious under 35 U.S.C. § 103.

118. I reserve the right to supplement by opinions in the future and to respond to any arguments that Petitioner or its expert may raise and to take into account any new information as it becomes available to me.

119. I have been warned that willful false statements and the like are punishable by fine or imprisonment, or both. I make this declaration of my own personal knowledge, and all statements made of my own knowledge are true. All

statements made on information and belief are believed to be true. If called to testify as to the truth of the matters stated herein, I could and would testify competently.

I declare under penalty of perjury that the foregoing is true and correct.

Executed on May 3, 2024 at Dover, Massachusetts.

A handwritten signature in black ink, appearing to read 'John Quackenbush', written over a horizontal line.

John Quackenbush, Ph.D.

APPENDIX A

CURRICULUM VITAE

Date: October 31, 2023

Name: John Quackenbush, Ph.D.

Office Address: Department of Biostatistics, 655 Huntington Avenue, Building II, Room 421, Boston, MA 02115

Home Address: 86 Walpole Street, Dover, MA 02030, USA

Work e-mail: johnq@hsph.harvard.edu

Work Phone: +1.617.432.9028

Work FAX +1.617.432.5619

Mobile Phone: +1.301.526.1740

Date & Place of Birth: January 4, 1962 Kingston, PA, USA

Education:

1983	Physics	B.S.	California Institute of Technology, Pasadena, CA
1984	Physics	M.S.	University of California, Los Angeles, CA
1990	Theoretical Particle Physics	Ph.D.	University of California, Los Angeles, CA

Postdoctoral Training:

1990-1992	Postdoctoral Fellow, Experimental Particle Physics, University of California, Los Angeles, CA
1992-1994	Staff Scientist, Molecular Genetics and Genomics, The Salk Institute for Biological Studies, La Jolla, CA
1994-1996	Research Associate, Genomics and Bioinformatics, Stanford Human Genome Center, Palo Alto, CA

Academic Appointments:

2018- present	Henry Pickering Walcott Professor of Computational Biology and Bioinformatics; Chair, Department of Biostatistics; Director, Harvard Health Data Science Center, Harvard TH Chan School of Public Health <u>Research Activities:</u> Development of methods in network medicine and systems biology aimed at understanding the multifactorial drivers of human health and disease, identifying disease subtypes and new therapeutic interventions, and exploring differences between the sexes in
---------------	--

- disease etiology and response to therapy; development of open-source software tools supporting research and clinical applications.
Activities: Founding Director/Co-Director, SM Program in Computational Biology and Quantitative Genetics (CBQG); Director, BD2K T32 Training Grant; Director T32 Cancer Training Grant; Founding Co-Director and Executive Committee, Joint Biostatistics and Bioinformatics T32 Training Grant; Director, Harvard Quantitative Biomedical Research Center.
- 2005- present Professor of Biostatistics and Computational Biology, Department of Data Science, Division of Computational Biology, Dana-Farber Cancer Institute;
Professor of Cancer Biology, Department of Cancer Biology, Dana-Farber Cancer Institute
Research Activities: Application of genomic technologies and integrative computational analysis to model cellular systems with an emphasis on understanding of human cancers and other diseases; development of open-source software tools supporting research and clinical applications.
- 2005- 2018 Professor of Computational Biology and Bioinformatics, Department of Biostatistics, Harvard TH Chan School of Public Health
Research Activities: Application of genomic technologies and integrative computational analysis to model cellular systems with an emphasis on understanding of human cancers and other diseases; development of open-source software tools supporting research and clinical applications.
Activities: Founding Director, SM Program in Computational Biology and Quantitative Genetics (CBQG); Director, BD2K T32 Training Grant; Founding Co-Director and Executive Committee, Joint Biostatistics and Bioinformatics T32 Training Grant.
- 2008- 2018 Director, Center for Cancer Computational Biology, Dana-Farber Cancer Institute
Research Activities: Integrative approaches to data analysis focused on clinical and translational applications and linking clinical and laboratory data in support of basic research; creation of software tools and systems to support research; overall analytical support for investigative research activities for all Dana-Farber Investigators.
- 2007 – present Adjunct Professor of Bioinformatics, Boston University
Activities: Mentoring PhD student thesis research.
- 2003- 2005 Professor, Chemical Engineering, University of Maryland
Activities: Mentoring PhD student thesis research.
- 2002 - 2006 Investigator, The Institute for Genomic Research
Research Activities: Development and implementation of technology and strategies, including computational methods and software tools,

necessary for functional analysis of the human and other genomes. Analysis of human gene expression in colon tumors using microarrays, rodent models of heart, lung, blood and sleep disorders and gene expression in Arabidopsis. Director of the TIGR Gene Index Project, providing freely available databases and software to estimate gene content in more than 100 organisms.

- 2000- 2005 Professor, Department of Biochemistry,
The George Washington University
Activities: Instructor in various courses, curriculum committee for Genomics.
- 1998 – 2005 Lecturer, The Department of Biostatistics,
The Johns Hopkins University
Activities: Instructor in the Masters In Biotechnology Program
- 1997 - 2001 Associate Investigator, The Institute for Genomic Research
Research Activities: Development and implementation of technology and strategies, including computational methods and software tools, necessary for functional analysis of the human and other genomes. Analysis of human gene expression in colon tumors using microarrays, rodent models of heart, lung, blood and sleep disorders and gene expression in Arabidopsis. Director of the TIGR Gene Index Project, providing freely available databases and software to estimate gene content in more than 50 organisms.
- 1997 Assistant Investigator, The Institute for Genomic Research
Research Activities: Development and implementation of technology and strategies, including analytical methods and software tools, necessary for functional analysis of the human and other genomes.
- 1994 - 1997 Research Associate, Stanford Human Genome Center
Stanford University
Research Activities: Project leader for development and implementation of a transposon-mediated strategy for large-scale genomic DNA sequencing including development and implementation of laboratory protocols, analytical methods, computer software, and instrumentation. Mapping, sequencing, and annotating regions of human chromosomes 4 and 21.
- 1992 - 1994 Staff Scientist, The Salk Institute for Biological Studies
Research Activities: Development of improved methods for DNA sequencing, combinatoric strategies and devices for screening large libraries, simulation and optimization of single pass sequencing strategies including genome sequence sampling. Project director for the STS content mapping of chromosome 11. Development of methods and software tools to support genome mapping.

- 1990 - 1992 Postdoctoral Fellow, Department of Physics
University of California, Los Angeles
Research Activities: Research in particle physics, field theory and phenomenology. Founding member of the Antiproton Experiment (APEX) collaboration that set the world's best experimental limit on the lifetime of the antiproton.
- 1987 - 1992 Visiting Lecturer, Department of Physics
University of California, Los Angeles
- 1984 - 1996 Physics Instructor, Southern California Science Institute
New College of California
- 1983 – 1990 Teaching Fellow, Department of Physics
University of California, Los Angeles
- 1983 - 1990 Research Fellow, Department of Physics
University of California, Los Angeles
Thesis Supervisor: E.T. Tomboulis
Research Activities: Research in elementary particle physics, mathematical physics, field theory focusing on the development of two-dimensional gauge field theory models and the construction of associated string theory models.

Nonacademic Employment

- 2011-2014 Co-Founder, CEO and Board Chairman, Genospace LLC
Activities: Genospace provides software solutions to operationalize precision medicine through the integration of clinical and genomic data and its presentation to clinical and research users in an intuitive format.
- 2014-2017 Co-Founder and Board Chairman, Genospace LLC
Activities: Genospace provides software solutions to operationalize precision medicine through the integration of clinical and genomic data and its presentation to clinical and research users in an intuitive format. In January 2017, Genospace was sold to the Hospital Corporation of America to support research and to acquire our world-leading solution for clinical trials matching.

Hospital or Affiliated Institution Appointments:

- 2003 – 2005 Adjunct Professor, Department of Biostatistics
Bloomberg School of Public Health, The Johns Hopkins University
- 2005- present Professor, Department of Biostatistics and Computational Biology
Dana-Farber Cancer Institute
- 2014- present Professor, Channing Division of Network Medicine, Brigham and Women's Hospital

Hospital and Health Care Organization Service Responsibilities:

- 2005- present Primary research areas are computational and system biology focused on the integration of diverse data types to provide insight into biological processes that drive human disease phenotypes.
Department of Data Science, Division of Computational Biology,
Dana-Farber Cancer Institute
- 2014- present Research at the interface of systems biology and pulmonary disease, exploration of how sex and gender affect disease risk, development of methods to better handle sex differences in systems approaches, and advising faculty in their research and the Director in setting directions for future work.
Channing Division of Network Medicine, Brigham and Women's Hospital

Major Committee Assignments:

- 1994 National Laboratory Genome Project Visit Review Panel, DOE
1995 Caltech/BAC Library Construction Site Visit Review Panel, DOE
1996-1998 Bioinformatics Grant Review Panel, DOE
1997-2001 Grant Review Panel, DOE
1997 Hollander Fellowship Review Panel, DOE
1997 Five-Year Program Advisory Committee, NCRR
1997 Cancer Chromosome Anatomy Project program Steering Committee, NCI
1998 National Laboratory Functional Genomic Review Panel, DOE
1998 Full-Length cDNA Library Construction and Sequencing Advisory, Committee, NCI
1998 Functional Genomics Panel, Wellcome Trust
1998 Genomics Grant Review Panel, NSF SBIR
1998 Special Program in Tropical Disease Research Review, UNDP/World Bank/WHO
1999-2000 Low-Dosage radiation Grant Review Panel, DOE
1999 Bioinformatics Review Panel, NSF
1999 Special Emphasis Review Panel, NHLBI
1999 Microarray Working Group Advisory Panel, NIDCR
1999 Special Emphasis Review Panel, NIMH
1999 Plant Biology Review Panel, NSF
1999-2007 Board of Directors, MGED
2000 Grant Review Panel, NIDA SBIR
2000 Plant Biology Review Panel, USDA
2000 Plant Biology Review Panel, NSF
2000 Exceptional Chromosome Regions Working Group, DOE
2000 Working Group on US Scientific Interactions, NSF
2000-2006 Genome/GCAT study section, NIH
2000-2003 PGA Coordinating Committee, HHLBI
2000-2003 PGA Bioinformatics Committee, NHLBI

2000-2003	PGA Microarray Committee, NHLBI
2000	Working Group on Australia Scientific Interactions, NSF
2001	Review Meeting RFP ES-01-01 National Center for Toxicogenomics (NCT) Microarray Resource, HIEHS
2001-2003	Plant Genome Scientific Advisory Board, University of Georgia
2001-2003	Innovative Technology Review Panel, NCI
2001-2002	Genome Study Section, NIH
2002-2006	GCAT Study Section, NIH
2001	Bioinformatics in Neuroscience and Sleep Research Workshop, NIDDK
2001	Plant Genome Site Visit Panel, NSF
2001-2003	BISTI Review Panel, NIH
2001-2005	Grant Review Panel, NCI
2002	Microarray Supplement Review Panel, NHLBI
2002-2006	Panel on Emerging Issues in Toxicogenomics, NRC
2004-2006	Panel on Applications of Toxicogenomics, NRC
2003-2010	Genome British Columbia Scientific Advisory Board
2005-2014	EPA Board of Scientific Counselors
2005-2011	St. Jude's Children's Research Hospital Scientific Advisory Board
2006-2007	Harvard School of Public Health, Lefkopoulou Award Committee Member
2006	Dana Farber Cancer Institute, Computational Biology Search Chair
2005	EU FP6 Grant Review Panel
2006-2007	Harvard School of Public Health, Seminar Committee Member
2006-2019	Harvard School of Public Health, Curriculum Committee Member
2006-2009	Harvard School of Public Health, Computational Biology Search Chair
2007-present	EU FP7 Genomics Grant Review Panel
2008-present	Lovelace Respiratory Research Institute, Scientific Advisory Board
2008-2010	NAS Panel on Collecting, Storing, and Distributing Biodata linked to Social Science Surveys
2009	RC1 Grant Review "Editorial Review" Panel
2009	RC2 Genomic Science Grant Review
2009	Epigenomics Roadmap Grant Review Panel
2009-2014	Epigenomics Roadmap Advisory Board
2010	NCI EUREKA Grant Review Panel
2010	Chair, NCRR S10 Grant Review Panel
2016	NASA ISS Genomic Applications Study Section
2019	NASA Human Research Program Long-Duration Flight Advisory Panel
2016	21 st Century Cures Act Biomarker Advisory Group
2012-2019	Research Oversight Committed, Genome British Columbia
2014-present	Shriner's Hospital standing review panel
2005-present	Various NIH grant review panels, either as an ad hoc member or as an ad hoc reviewer. I typically serve on at least six per year and can provide a more detailed list on request.
2005-present	Various EU and Canadian grant review panels, either as an ad hoc member or as an ad hoc reviewer. I can assemble a detailed list on request.
2017-present	Research Foundation Flanders (FWO) standing review panel.
2018-present	Chair, Department of Biostatistics, Harvard T.H. Chan School of Public Health

Nonacademic Board and Advisory Board Positions (active or past five years):

2005-2016	Nabsys Inc., Scientific Advisory Board
2010-2017	Lovelace Respiratory Research Institute, Scientific Advisory Board
2012-2016	SynapDx, Scientific Advisory Board
2012-2017	Genospace LLC, Founder, CEO, Board of Directors; Chairman of the Board
2012-2020	Blueprint Genomics, Scientific Advisory Board
2013-2017	Thomson-Reuters, Genomics Advisory Board
2013-present	Caris Life Sciences, Scientific Advisory Board
2020-present	Caris Life Sciences, Data Science Advisory Board
2014-present	Shriner's Hospitals, Scientific Advisory Board
2014-present	Olema Pharmaceuticals, Scientific Advisory Board
2015-present	Haymakers for Hope, Board of Directors
2015-present	Lacriscience, Scientific Advisory Board
2017-present	Merck KGaA, Precision Medicine Advisory Board
2018-present	Renalytix AI, Advisory Board
2018-present	Dover, MA, Capital Budget Committee
2019-present	StitchBio, Scientific Advisory Board
2020-present	Charles River School Health and Safety Committee (COVID Advisory Board)
2020-present	Boston University Academy Health and Safety Committee (COVID Advisory Board)
2021-present	Dover, MA, Board of Health Special Advisor
2021-present	Congreso Futuro International Advisory Board, Chile

Professional Societies:

1988	American Physical Society	Member
1988	APS Biophysics Division	Member
1988	APS International Physics Group	Member
1988	APS Division of Particles and Fields	Member
1988	World Federation of Scientists	Member
1990	Microarray Gene Expression Data Society (MGED)/ Functional Genomics Data Society (FGED)	Member Board Member
	(2011-2014)	President
1995	The Human Genome Organization (HUGO)	Member
1996	Association of Science Professionals (ASP)	Member
1998	American Association for the Advancement of Science (AAAS)	Member
1999	Center for the Study of the Evolution of Life (CSEOL)	Member
2005	American Association for Cancer Research	Member

Community Service Related to Professional Work:

I have presented more than 350 talks, courses, and workshops since joining Dana-Farber and the Harvard T.H. Chan School of Public Health in 2005; detailed lists will be

provided upon request. I have also taught numerous courses and workshops on the analysis of genomic data and have served as an organizer of the EMBO course on DNA microarray analysis for more than five years.

In addition, I have served on the Scientific Advisory Board of the St. Jude's Children's Research Hospital, the Scientific Advisory Board for the Lovelace Respiratory Research Institute, Research Oversight Committees for various Genome Canada programs, and the Board of Scientific Counselors of the Environmental Protection Agency. I have also been a member of advisory boards for a number of NIH initiatives, including the Roadmap Epigenome Project, the Extracellular RNA Communication Consortium, the Atopic Dermatitis Research Network (ADRN), and the Trans-Omics for Precision Medicine (TOPMed) program.

Since the onset of the COVID19 pandemic, I have served numerous local school and community groups by advising these on policy and practice. I am particularly proud that the two schools I advised instituted a series of policies addressing air quality, masking, distancing, hygiene, and vaccination that allowed them to remain open without any documented cases of in-school transmission during the 2020-2021 and 2021-2022 academic years. I have also become a special advisor to the Dover, MA Board of Health and provide advice to the Dover-Sherborn Regional School District on COVID and other health-related matters.

(Peer Review for Scientific Journals)

1994-2016	<i>Genomics</i>
1994-2009	<i>BioTechniques</i>
1996-2010	<i>Genome Research</i>
2000-2010	<i>Genome Biology</i>
1997-1998	<i>Analytical Biochemistry</i>
1997-present	<i>Nature</i>
1997-2010	<i>Nucleic Acids Research</i>
1999-2010	<i>Nature Genetics</i>
1999-2009	<i>EMBO Reports</i>
1999-2017	<i>Nature Biotechnology</i>
1999-present	<i>Science</i>
2000-2016	<i>Bioinformatics</i>
2000-2009	<i>Journal of Theoretical Biology</i>
2000-2009	<i>Mammalian Genomics</i>
2000-2006	<i>Gene</i>
2000-2016	<i>BMC Genomics, BMC Bioinformatics</i>
2001-2009	<i>Structural and Functional Genomics</i>
2001-2010	<i>Nature Reviews Genetics</i>
2001-2006	<i>American Journal of Applied Statistics</i>
2009-present	<i>PLoS Computational Biology</i>
2015-present	<i>Proceedings of the National Academy of Sciences</i>
2017-present	<i>Molecular Systems Biology</i>

Editorial Boards:

2000-2009	Editorial Board	<i>BioTechniques</i>
2000-2012	Associate Editor	<i>Bioinformatics</i>
2000-present	Editorial Board	<i>BMC Genomics</i>
2000-2006	Editorial Board	<i>Genomics</i>
2006-2017	Editor-in-Chief	<i>Genomics</i>
2018-present	Editorial Board	<i>Cancer Research</i>
2022-present	Editorial Board	<i>GEN Biotechnology</i>

Awards and Honors:

1985 - 1991	Outstanding Teaching Award (7) , UCLA Physics Department
1986	Jun John Sakurai Scholarship , 24 th International School of Subnuclear Physics, Ettore Majorana Centre for Scientific Culture, Erice, Italy
1987	UCLA Distinguished Teaching Assistant Award , UCLA Academic Senate
1987	Prize for Best Student and Prize for Best Scientific Secretary , 25 th International School of Subnuclear Physics, Ettore Majorana Centre for Scientific Culture, Erice, Italy
1988	Graduate Distinguished Scholar Award , UCLA Alumni Association
1992 - 1997	SERCA Fellow , National Human Genome Research Institute
2006	George D. Wilbanks Lectureship in Gynecological Oncology , University of South Florida College of Medicine
2007	President's Distinguished Lectureship , American Society of Reproductive Medicine
2007	Leopold G. Koss Lectureship , Universität Bern, Switzerland
2007	Distinguished International Advisors at IEE BIBE, the original and the oldest IEEE flagship international conference in Bioinformatics and Bioengineering.
2008	Distinguished Lecture in Computer Science , Wayne State University
2010	IPM Distinguished Lecture , Mt. Sinai Medical School, New York
2010	The Ian Lawson Van Toch Memorial Seminar Series in Computational Biology , Ontario Cancer Institute
2010	Harvard-Australia Foundation Fellowship
2010	Bancroft Fellow-in-Residence , Queensland Institute for Medical Research, Queensland, Australia
2010	Jackson Memorial Fellowship , Griffith University, Queensland, Australia
2011	Australia Fellow , National Health and Medical Research Council, Australia
2013	White House Open Science Champion of Change , Office of the President of the United States of America
2016	Mentor of the Year Award , Postdoc and Graduate Student Organization, Dana-Farber Cancer Institute
2018	National Cancer Institute Outstanding Investigator Award (R35)

2022

Elected Fellow, National Academy of Medicine

Part II: Research, Teaching, and Clinical Contributions**A. Narrative report (700 words or less) of Research, Teaching, and Clinical Contributions. Please focus primarily on the areas in which most of your time and effort is spent.**

My current research involves development of new analytical methods for the integrated analysis of large-scale biomedical data, including genomic data, to better understand the nature of human diseases, including cancer and pulmonary diseases, with the goal of developing better methods for diagnosis and treatment. Two defining characteristics of my work are a commitment to reproducible research through development of open-source software implementations of analytical methods, and an emphasis on studying sexual dimorphism in health and disease.

A longstanding component of my work is methods development for robust biomarkers. We have applied this to identifying reproducible predictive biomarkers in cancer and have explored reproducibility in large, published data sets, demonstrating that many are inconsistent. In collaboration with Hugo Aerts, I have been working to extend our methods to the analysis of quantitative feature data extracted from radiographic images (an emerging field called “radiomics”), to use these to identify predictive biomarkers, and to link these features to genetic mutations in cancer. I have also come to recognize that while individual “features” such as the expression of genes or specific mutations may not be sufficient on their own to identify drivers of biological states or to predict outcome or response to therapy in disease.

I have also been developing methods for inferring gene regulatory network models using gene expression and other sources of data. We published a method, PANDA, that builds on “message passing” (an idea developed communications theory) to optimize information flow through a gene regulatory network model, allowing us to infer condition-specific networks and to compare these between phenotypes. We have developed many extensions of PANDA that incorporate other data types and have applied these to the study of diseases, including ovarian cancer, colorectal cancer, chronic obstructive pulmonary disease, asthma, and Alzheimer’s disease, as well as to the investigation of sexual dimorphism. Consistent with our commitment to open and reproducible science, we have collected these methods (all named for animals) with standardized source code in R, C, Matlab, and Python, into the “Network Zoo” (<https://netzoo.github.io/>). To support the widespread use of these methods, we developed “netBooks” (<http://netbooks.networkmedicine.org/>), a collection of online tutorials that use Jupyter notebooks to introduce the use of our methods and, increasingly, to allow all of our published analyses to be rerun by those interested in the analyses.

We have also been exploring the connection between genotype and phenotype, using gene expression as a surrogate for phenotype. We have developed a new approach, CONDOR, to the analysis of expression quantitative trait loci that uses a bi-partite graph to represent associations between gene expression levels and both local (*cis-acting*) and

distant (*trans-acting*) significant genetic variants. We find that the overall structure of the graph provides insight not only into how genetic variants work together to influence phenotype, but also into evolutionary constraints on the genetic variants themselves. This has given us new insight into how many genetic variants of small effect-size can collectively exert strong influence on phenotype.

More recently, we have been asking whether how one can model regulatory networks in the context of genetic background, recognizing that an individual's genotype will manifest its effects through alterations of patterns of gene expression that ultimately will influence the functional processes active in individuals. That method, EGRET, represents an important step forward in the analysis of the link between genotype and phenotype and is the first method for linking genetic cause and phenotypic effect through the action of individual regulatory processes.

Our group also builds and maintains a variety of software tools to support genomic analysis. the TM4 suite of open-source software tools for the analysis of DNA microarray data. The most widely used of these, MeV, is a powerful data mining software tool, that places sophisticated analytical tools behind an intuitive graphical user interface. Developed to be accessible to laboratory biologists, MeV has more than 260,000 users worldwide. We recently transitioned MeV to a cloud-based platform, WebMeV (<http://mev.tm4.org>) to take advantage of cloud scalability to better handle large sequence-based data. WebMeV now has 5000 registered users and provides more than 200 analysis sessions per day.

I am also very involved in teaching and mentoring. I developed, launched, and direct the Computational Biology and Quantitative Genetics (CBQG) master's degree program, I was co-PI and co-director of the first joint Biostatistics and Bioinformatics training grant in the US and recently stepped down to develop and direct a training grant program as part of the NIH's Big Data to Knowledge (BD2K) program. I also developed and teach an introductory genomics tools and methods course and have co-taught an online course on reproducible research through HarvardX (<https://www.edx.org/course/principles-statistical-computational-harvardx-ph527x>).

I have trained more than 80 postdoctoral fellows and graduate students (including MS and PhD students whom I have mentored for their thesis projects) and have mentored a number of junior faculty members since joining the Harvard TH Chan School of Public Health and Dana-Farber Cancer Institute in 2005. I have emphasized diversity in hiring postdocs; more than half of my trainees have been female and four of the five postdocs from my group who have moved onto faculty positions in the past year have been female. I also maintain an extensive network of former trainees (and their current trainees) which meets weekly as we continue to work collaboratively in the development and application of network methods.

B. Funding Information

I have had more than twenty-five years of continuous grant support from the NIH and other organizations. My current support includes:

Active Research Support**5R01CA205406-03 (Ng, K.)****3/1/2017 – 2/28/2022**

NIH/NCI

Novel Randomized Controlled Trials of Vitamin D Supplementation in Patients with Colorectal Cancer: Impact on Survival and Biology

This project is exploring the biology underlying the effects of vitamin D on survival of colorectal patients.

Role: Co-Investigator

1R35CA220523 (Quackenbush, J.)**9/1/2018 – 8/31/2025**

NIH/NCI

Unraveling the Complexities of Risk and Mechanism in Cancer

This NCI Career Award supports Dr. Quackenbush's ongoing work to model genetic processes in cancer and to elucidate the link between genotype and phenotype so as to better understand the drivers of cancer and to identify potential targets for therapeutic intervention, with a particular emphasis on understanding differences between men and women. Role: Principal Investigator

1R01HG011393 (DeMeo, D / Quackenbush, J.)**9/1/2018 – 06/30/2025**

NIH/NCI

Network Tools to Understand Sex- and Gender-Specific Drivers of Disease

This project will develop robust methods for handling male and female multi-omic data, with an emphasis on applying these to the inference of gene regulatory network models and the use of such models for gaining insight into the regulatory processes that drive clinical differences in human health and disease. Role: co-Principal Investigator

1U24CA231846-01A1 (Quackenbush, J.)**9/1/2019 – 8/31/2024**

NIH/NCI

MeV: A Robust Platform for Intuitive Genomic Data Analysis

As part of the Informatics Technology for Cancer Research (ITCR) program, we will develop and support WebMeV, a robust, scalable data analysis software tool that uses intuitive visual interfaces to provide users with access to advanced data analysis methods for large-scale genomic data in a system that helps ensure their research is reproducible.

Role: Principal Investigator

2P50CA127003-11A1 (Bass, A.)**9/1/2019 – 5/31/2024**

NIH (Subcontract: Dana Farber Cancer Institute)

Dana Farber/ Harvard Cancer Center SPORE in Gastrointestinal Cancer

The GI SPORE supports state-of-the-art investigator-initiated translational research that will contribute to improved prevention, early detection, diagnosis, and treatment of gastrointestinal cancers. Dr. Quackenbush will serve as Co-Director of the Biostatistics and Bioinformatics Core, which will support the quantitative needs of the investigators in the design of experiments and the analysis of the resulting data.

Role: Co-Investigator/Core Co-Director

C. Report of Current Research Activities other than those mentioned above (bench research, clinical trials, outcome studies, efficacy studies as applicable)

None to report.

D. Report of Teaching (use only those categories that are applicable)

1983 - 1990	Teaching Fellow, Department of Physics, University of California, Los Angeles
1984 - 1996	Physics Instructor, Southern California Science Institute, New College of California
1987 - 1992	Visiting Lecturer, Department of Physics, University of California, Los Angeles
1998 - 2005	Lecturer, The Johns Hopkins University Instructor in the Masters in Biotechnology Program
2000 - 2005	Professor, Department of Biochemistry, The George Washington University
2003 - 2005	Adjunct Professor, Department of Biostatistics, Bloomberg School of Public Health, The Johns Hopkins University Developed program in genomics and bioinformatics; instructor in various courses.
2005-present	Professor, Department of Biostatistics, Harvard TH Chan School of Public Health
2013-present	Director, MS in Computational Biology and Quantitative Genetics

E. Report of Clinical Activities

1. Description of clinical practice (field, areas of major focus, site(s) of practice [private office, HMO, teaching hospital etc.])

None

2. Patient load (indicate complexity of cases, as appropriate)

None

3. Clinical contributions (e.g., introduction of new methods of clinical diagnosis, prevention, treatment, care delivery)

None

4. Other relevant information about clinical role (receipt of clinical awards, locally or nationally, invitation to participate in clinical activities at other sites, special recognition by peers or professional organizations as a leader in a clinical field)

None

Part III: Bibliography (as of October 31, 2023)

Total citations: 93,116

h-index: 110

i10-index: 308

Original Articles (325)

Complete List of Published Works in MyBibliography:

<https://www.ncbi.nlm.nih.gov/myncbi/john.quackenbush.1/bibliography/public/>

(Note that this list includes publications by students funded on training grants on which I am a PI.)

1. **Quackenbush J.** Chiral anomalies in two-dimensional quantum field theory. Physical review D: Particles and fields. 1989;40(10):3408-14. Epub 1989/11/15. PubMed PMID: 10011709.
2. **Quackenbush J.** Twisted world-sheet gauge fields and string models Phys Lett B. 1990;234(3):285-91.
3. Buchanan CD, Cousins R, Dib C, Peccei RD, **Quackenbush J.** Testing CP and CPT violation in the neutral kaon system at a phi factory. Physical review D: Particles and fields. 1992;45(11):4088-107. Epub 1992/06/01. PubMed PMID: 10014315.
4. Geer S, Marriner J, Ray R, Streets J, Lindgren M, Muller T, **Quackenbush J**, Armstrong T. Search for antiproton decay at the Fermilab antiproton accumulator. Physical review letters. 1994;72(11):1596-9. Epub 1994/03/14. PubMed PMID: 10055652.
5. Khristich JV, Bailis J, Diggie K, Rodkins A, Romo A, **Quackenbush J**, Evans GA. Large-scale screening of yeast artificial chromosome libraries using PCR. BioTechniques. 1994;17(3):498-501. Epub 1994/09/01. PubMed PMID: 7818903.
6. **Quackenbush J.** Career issues. Science. 1994;266(5189):1306. Epub 1994/11/25. doi: 10.1126/science.266.5189.1306. PubMed PMID: 17772824.
7. Selleri L, Giovannini M, Hermanson GG, Romo A, **Quackenbush J**, Penny L, Khristich JV, Evans GA. Yeast artificial chromosome cloning of 3.2 megabases within chromosomal band 11q24 closely linking c-ets 1 and Fli-1 and encompassing the Ewing sarcoma breakpoint. Genomics. 1994;22(1):137-47. Epub 1994/07/01. doi: 10.1006/geno.1994.1354. PubMed PMID: 7959760.
8. Kupfer K, Smith MW, **Quackenbush J**, Evans GA. Physical mapping of complex genomes by sampled sequencing: a theoretical analysis. Genomics. 1995;27(1):90-100. Epub 1995/05/01. doi: 10.1006/geno.1995.1010. PubMed PMID: 7665188.
9. **Quackenbush J**, Davies C, Bailis JM, Khristich JV, Diggie K, Marchuck Y, Tobin J, Clark SP, Rodkins A, Marcano S, et al. An STS content map of human chromosome 11: localization of 910 YAC clones and 109 islands. Genomics. 1995;29(2):512-25. Epub 1995/09/20. PubMed PMID: 8666402.
10. Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tome P, Aggarwal A, Bajorek E, Bentolila S, Birren BB, Butler A, Castle AB, Chiannikulchai N, Chu A, Clee C, Cowles S, Day PJ, Dibling T, Drouot N, Dunham I, Duprat S, East C, Edwards C, Fan JB, Fang N, Fizames C,

- Garrett C, Green L, Hadley D, Harris M, Harrison P, Brady S, Hicks A, Holloway E, Hui L, Hussain S, Louis-Dit-Sully C, Ma J, MacGilvery A, Mader C, Maratukulam A, Matise TC, McKusick KB, Morissette J, Mungall A, Muselet D, Nusbaum HC, Page DC, Peck A, Perkins S, Piercy M, Qin F, **Quackenbush J**, Ranby S, Reif T, Rozen S, Sanders C, She X, Silva J, Slonim DK, Soderlund C, Sun WL, Tabar P, Thangarajah T, Vega-Czarny N, Vollrath D, Voyticky S, Wilmer T, Wu X, Adams MD, Auffray C, Walter NA, Brandon R, Dehejia A, Goodfellow PN, Houlgatte R, Hudson JR, Jr., Ide SE, Iorio KR, Lee WY, Seki N, Nagase T, Ishikawa K, Nomura N, Phillips C, Polymeropoulos MH, Sandusky M, Schmitt K, Berry R, Swanson K, Torres R, Venter JC, Sikela JM, Beckmann JS, Weissenbach J, Myers RM, Cox DR, James MR, Bentley D, Deloukas P, Lander ES, Hudson TJ. A gene map of the human genome. *Science*. 1996;274(5287):540-6. Epub 1996/10/25. PubMed PMID: 8849440.
11. Doyle DJ, **Quackenbush J**. Symposium on Genomic Medicine, University of Maryland, Shady Grove Campus, Rockville, Maryland, March 17-18, 1997. *Microbial & comparative genomics*. 1997;2(2):99-102. Epub 1997/01/01. PubMed PMID: 9689218.
 12. Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK, Gwinn M, Dougherty B, Tomb JF, Fleischmann RD, Richardson D, Peterson J, Kerlavage AR, **Quackenbush J**, Salzberg S, Hanson M, van Vugt R, Palmer N, Adams MD, Gocayne J, Weidman J, Utterback T, Wathley L, McDonald L, Artiach P, Bowman C, Garland S, Fujii C, Cotton MD, Horst K, Roberts K, Hatch B, Smith HO, Venter JC. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature*. 1997;390(6660):580-6. Epub 1997/12/24. doi: 10.1038/37551. PubMed PMID: 9403685.
 13. Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, **Quackenbush J**, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA, McKenney K, Adams MD, Loftus B, Peterson S, Reich CI, McNeil LK, Badger JH, Glodek A, Zhou L, Overbeek R, Gocayne JD, Weidman JF, McDonald L, Utterback T, Cotton MD, Spriggs T, Artiach P, Kaine BP, Sykes SM, Sadow PW, D'Andrea KP, Bowman C, Fujii C, Garland SA, Mason TM, Olsen GJ, Fraser CM, Smith HO, Woese CR, Venter JC. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature*. 1997;390(6658):364-70. Epub 1997/12/06. doi: 10.1038/37052. PubMed PMID: 9389475.
 14. Korenberg JR, Aaltonen J, Brahe C, Cabin D, Creau N, Delabar JM, Doering J, Gardiner K, Hubert RS, Ives J, Kessler A, Kudoh J, Lafreniere R, Murakami Y, Ohira M, Ohki M, Patterson D, Potier MC, **Quackenbush J**, Reeves RH, Sakaki Y, Shimizu N, Soeda E, Van Broeckhoven C, Yaspo ML. Report and abstracts of the Sixth International Workshop on Human Chromosome 21 Mapping 1996. Cold Spring Harbor, New York, USA. May 6-8, 1996. *Cytogenetics and cell genetics*. 1997;79(1-2):21-52. Epub 1997/01/01. PubMed PMID: 9533011.
 15. Stewart EA, McKusick KB, Aggarwal A, Bajorek E, Brady S, Chu A, Fang N, Hadley D, Harris M, Hussain S, Lee R, Maratukulam A, O'Connor K, Perkins S, Piercy M, Qin F, Reif T, Sanders C, She X, Sun WL, Tabar P, Voyticky S, Cowles S, Fan JB, Mader C, **Quackenbush J**, Myers RM, Cox DR. An STS-based

- radiation hybrid map of the human genome. *Genome research*. 1997;7(5):422-33. Epub 1997/05/01. PubMed PMID: 9149939.
16. Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, **Quackenbush J**, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzgerald LM, Lee N, Adams MD, Hickey EK, Berg DE, Gocayne JD, Utterback TR, Peterson JD, Kelley JM, Cotton MD, Weidman JM, Fujii C, Bowman C, Watthey L, Wallin E, Hayes WS, Borodovsky M, Karp PD, Smith HO, Fraser CM, Venter JC. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature*. 1997;388(6642):539-47. Epub 1997/08/07. doi: 10.1038/41483. PubMed PMID: 9252185.
 17. **Quackenbush J**, Adams MD, Fraser CM, Venter JC. Genome Projects: From Microbes to Man, From Sequence to Function. In: Dixon GK, Copping LG, Livingstone D, editors. *Genomics: Commercial Opportunities from a Scientific Revolution*: BIOS Scientific Publishers; 1998.
 18. Kelly JM, **Quackenbush J**. Sequencing PCR Products. In: Innis M, Gelfand D, Sninsky J, editors. *PCR Methods Manual*: Academic Press; 1999.
 19. Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*. 2000;408:796-815.
 20. El-Sayed NM, Hegde P, **Quackenbush J**, Melville SE, Donelson JE. The African trypanosome genome. *International journal for parasitology*. 2000;30(4):329-45. Epub 2000/03/25. PubMed PMID: 10731558.
 21. Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snedrud E, Lee N, **Quackenbush J**. A concise guide to cDNA microarray analysis. *BioTechniques*. 2000;29(3):548-50, 52-4, 56 passim. Epub 2000/09/21. PubMed PMID: 10997270.
 22. Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, **Quackenbush J**. Gene index analysis of the human genome estimates approximately 120,000 genes. *Nature genetics*. 2000;25(2):239-40. Epub 2000/06/03. doi: 10.1038/76126. PubMed PMID: 10835646.
 23. Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, **Quackenbush J**. An optimized protocol for analysis of EST sequences. *Nucleic acids research*. 2000;28(18):3657-65. Epub 2000/09/13. PubMed PMID: 10982889; PubMed Central PMCID: PMC110731.
 24. **Quackenbush J**. Viva la revolution! A report from the FANTOM meeting. *Nature genetics*. 2000;26(3):255-6. Epub 2000/11/04. doi: 10.1038/81525. PubMed PMID: 11062453.
 25. **Quackenbush J**, Liang F, Holt I, Pertea G, Upton J. The TIGR gene indices: reconstruction and representation of expressed gene sequences. *Nucleic acids research*. 2000;28(1):141-5. Epub 1999/12/11. PubMed PMID: 10592205; PubMed Central PMCID: PMC102391.
 26. Yuan Q, Liang F, Hsiao J, Zismann V, Benito MI, **Quackenbush J**, Wing R, Buell R. Anchoring of rice BAC clones to the rice genetic map in silico. *Nucleic acids research*. 2000;28(18):3636-41. Epub 2000/09/13. PubMed PMID: 10982886; PubMed Central PMCID: PMC110739.
 27. Brazma A, Hingamp P, **Quackenbush J**, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-

- Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature genetics*. 2001;29(4):365-71. Epub 2001/12/01. doi: 10.1038/ng1201-365. PubMed PMID: 11726920.
28. Flores-Morales A, Stahlberg N, Tollet-Egnell P, Lundeberg J, Malek RL, **Quackenbush J**, Lee NH, Norstedt G. Microarray analysis of the in vivo effects of hypophysectomy and growth hormone treatment on gene expression in the rat. *Endocrinology*. 2001;142(7):3163-76. Epub 2001/06/21. doi: 10.1210/endo.142.7.8235. PubMed PMID: 11416039.
 29. Gaspard R, Dharap S, Malek J, Qi R, **Quackenbush J**. Optimized growth conditions for direct amplification of cDNA clone inserts from culture. *BioTechniques*. 2001;31(1):35-6. Epub 2001/07/24. PubMed PMID: 11464517.
 30. Hegde P, Qi R, Gaspard R, Abernathy K, Dharap S, Earle-Hughes J, Gay C, Nwokekeh NU, Chen T, Saeed AI, Sharov V, Lee NH, Yeatman TJ, **Quackenbush J**. Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer research*. 2001;61(21):7792-7. Epub 2001/11/03. PubMed PMID: 11691794.
 31. Huang J, Qi R, **Quackenbush J**, Dauway E, Lazaridis E, Yeatman T. Effects of ischemia on gene expression. *The Journal of surgical research*. 2001;99(2):222-7. Epub 2001/07/27. doi: 10.1006/jsre.2001.6195. PubMed PMID: 11469890.
 32. Kappe SH, Gardner MJ, Brown SM, Ross J, Matuschewski K, Ribeiro JM, Adams JH, **Quackenbush J**, Cho J, Carucci DJ, Hoffman SL, Nussenzweig V. Exploring the transcriptome of the malaria sporozoite stage. *Proceedings of the National Academy of Sciences of the United States of America*. 2001;98(17):9895-900. Epub 2001/08/09. doi: 10.1073/pnas.171185198. PubMed PMID: 11493695; PubMed Central PMCID: PMC55549.
 33. Kawai J, Shinagawa A, Shibata K, Yoshino M, Itoh M, Ishii Y, Arakawa T, Hara A, Fukunishi Y, Konno H, Adachi J, Fukuda S, Aizawa K, Izawa M, Nishi K, Kiyosawa H, Kondo S, Yamanaka I, Saito T, Okazaki Y, Gojobori T, Bono H, Kasukawa T, Saito R, Kadota K, Matsuda H, Ashburner M, Batalov S, Casavant T, Fleischmann W, Gaasterland T, Gissi C, King B, Kochiwa H, Kuehl P, Lewis S, Matsuo Y, Nikaido I, Pesole G, **Quackenbush J**, Schriml LM, Staubli F, Suzuki R, Tomita M, Wagner L, Washio T, Sakai K, Okido T, Furuno M, Aono H, Baldarelli R, Barsh G, Blake J, Boffelli D, Bojunga N, Carninci P, de Bonaldo MF, Brownstein MJ, Bult C, Fletcher C, Fujita M, Gariboldi M, Gustincich S, Hill D, Hofmann M, Hume DA, Kamiya M, Lee NH, Lyons P, Marchionni L, Mashima J, Mazzarelli J, Mombaerts P, Nordone P, Ring B, Ringwald M, Rodriguez I, Sakamoto N, Sasaki H, Sato K, Schonbach C, Seya T, Shibata Y, Storch KF, Suzuki H, Toyo-oka K, Wang KH, Weitz C, Whittaker C, Wilming L, Wynshaw-Boris A, Yoshida K, Hasegawa Y, Kawaji H, Kohtsuki S, Hayashizaki Y, Team RGERGPI, the FC. Functional annotation of a full-length mouse cDNA collection. *Nature*. 2001;409(6821):685-90. Epub 2001/02/24. doi: 10.1038/35055500. PubMed PMID: 11217851.
 34. **Quackenbush J**. Crop Plant Genome Yield Huge Prospects. In: Butler D, editor. *The Nature Yearbook of Science and Technology 2001* London, UK: Palgrave Macmillan; 2001.

35. **Quackenbush J.** Computational analysis of microarray data. *Nature reviews Genetics*. 2001;2(6):418-27. Epub 2001/06/05. doi: 10.1038/35076576. PubMed PMID: 11389458.
36. **Quackenbush J.** The power of public access: the human genome project and the scientific process. *Nature genetics*. 2001;29(1):4-6. Epub 2001/08/31. doi: 10.1038/ng0901-4. PubMed PMID: 11528377.
37. **Quackenbush J**, Cho J, Lee D, Liang F, Holt I, Karamycheva S, Parvizi B, Pertea G, Sultana R, White J. The TIGR Gene Indices: analysis of gene transcript sequences in highly sampled eukaryotic species. *Nucleic acids research*. 2001;29(1):159-64. Epub 2000/01/11. PubMed PMID: 11125077; PubMed Central PMCID: PMC29813.
38. 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 research*. 2001;11(4):626-30. Epub 2001/04/03. doi: 10.1101/gr.170101. PubMed PMID: 11282978; PubMed Central PMCID: PMC311058.
39. Tsai J, Sultana R, Lee Y, Pertea G, Karamycheva S, Antonescu V, Cho J, Parvizi B, Cheung F, **Quackenbush J**. RESOURCERER: a database for annotating and linking microarray resources within and across species. *Genome biology*. 2001;2(11):SOFTWARE0002. Epub 2005/09/22. PubMed PMID: 16173164; PubMed Central PMCID: PMC138985.
40. Yuan Q, **Quackenbush J**, Sultana R, Pertea M, Salzberg SL, Buell CR. Rice bioinformatics. analysis of rice sequence data and leveraging the data to other plant species. *Plant physiology*. 2001;125(3):1166-74. Epub 2001/03/13. PubMed PMID: 11244096; PubMed Central PMCID: PMC1539370.
41. Agrawal D, Chen T, Irby R, **Quackenbush J**, Chambers AF, Szabo M, Cantor A, Coppola D, Yeatman TJ. Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. *Journal of the National Cancer Institute*. 2002;94(7):513-21. Epub 2002/04/04. PubMed PMID: 11929952.
42. Andersson T, Unneberg P, Nilsson P, Odeberg J, **Quackenbush J**, Lundberg J. Monitoring of representational difference analysis subtraction procedures by global microarrays. *BioTechniques*. 2002;32(6):1348-50, 52, 54-6, 58. Epub 2002/06/21. PubMed PMID: 12074166.
43. Ball CA, Sherlock G, Parkinson H, Rocca-Sera P, Brooksbank C, Causton HC, Cavalieri D, Gaasterland T, Hingamp P, Holstege F, Ringwald M, Spellman P, Stoeckert CJ, Jr., Stewart JE, Taylor R, Brazma A, **Quackenbush J**, Microarray Gene Expression D. The underlying principles of scientific publication. *Bioinformatics*. 2002;18(11):1409. Epub 2002/11/09. PubMed PMID: 12424109.
44. Ball CA, Sherlock G, Parkinson H, Rocca-Sera P, Brooksbank C, Causton HC, Cavalieri D, Gaasterland T, Hingamp P, Holstege F, Ringwald M, Spellman P, Stoeckert CJ, Jr., Stewart JE, Taylor R, Brazma A, **Quackenbush J**, Microarray Gene Expression Data S. Standards for microarray data. *Science*. 2002;298(5593):539. Epub 2002/10/22. PubMed PMID: 12387284.
45. Carlton JM, Angiuoli SV, Suh BB, Kooij TW, Pertea M, Silva JC, Ermolaeva MD, Allen JE, Selengut JD, Koo HL, Peterson JD, Pop M, Kosack DS, Shumway MF, Bidwell SL, Shallom SJ, van Aken SE, Riedmuller SB, Feldblyum TV, Cho JK,

- Quackenbush J**, Sedegah M, Shoaibi A, Cummings LM, Florens L, Yates JR, Raine JD, Sinden RE, Harris MA, Cunningham DA, Preiser PR, Bergman LW, Vaidya AB, van Lin LH, Janse CJ, Waters AP, Smith HO, White OR, Salzberg SL, Venter JC, Fraser CM, Hoffman SL, Gardner MJ, Carucci DJ. Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature*. 2002;419(6906):512-9. Epub 2002/10/09. doi: 10.1038/nature01099. PubMed PMID: 12368865.
46. Fahrenkrug SC, Smith TP, Freking BA, Cho J, White J, Vallet J, Wise T, Rohrer G, Perteu G, Sultana R, **Quackenbush J**, Keele JW. Porcine gene discovery by normalized cDNA-library sequencing and EST cluster assembly. *Mammalian genome : official journal of the International Mammalian Genome Society*. 2002;13(8):475-8. Epub 2002/09/13. doi: 10.1007/s00335-001-2072-4. PubMed PMID: 12226715.
 47. Kim H, Zhao B, Snesrud EC, Haas BJ, Town CD, **Quackenbush J**. Use of RNA and genomic DNA references for inferred comparisons in DNA microarray analyses. *BioTechniques*. 2002;33(4):924-30. Epub 2002/10/26. PubMed PMID: 12398202.
 48. Lee Y, Sultana R, Perteu G, Cho J, Karamycheva S, Tsai J, Parvizi B, Cheung F, Antonescu V, White J, Holt I, Liang F, **Quackenbush J**. Cross-referencing eukaryotic genomes: TIGR Orthologous Gene Alignments (TOGA). *Genome research*. 2002;12(3):493-502. Epub 2002/03/05. doi: 10.1101/gr.212002. PubMed PMID: 11875039; PubMed Central PMCID: PMC155294.
 49. Malek RL, Irby RB, Guo QM, Lee K, Wong S, He M, Tsai J, Frank B, Liu ET, **Quackenbush J**, Jove R, Yeatman TJ, Lee NH. Identification of Src transformation fingerprint in human colon cancer. *Oncogene*. 2002;21(47):7256-65. Epub 2002/10/09. doi: 10.1038/sj.onc.1205900. PubMed PMID: 12370817.
 50. Nene V, Lee D, **Quackenbush J**, Skilton R, Mwaura S, Gardner MJ, Bishop R. AvGI, an index of genes transcribed in the salivary glands of the ixodid tick *Amblyomma variegatum*. *International journal for parasitology*. 2002;32(12):1447-56. Epub 2002/10/24. PubMed PMID: 12392910.
 51. Okazaki Y, Furuno M, Kasukawa T, Adachi J, Bono H, Kondo S, Nikaido I, Osato N, Saito R, Suzuki H, Yamanaka I, Kiyosawa H, Yagi K, Tomaru Y, Hasegawa Y, Nogami A, Schonbach C, Gojobori T, Baldarelli R, Hill DP, Bult C, Hume DA, **Quackenbush J**, Schriml LM, Kanapin A, Matsuda H, Batalov S, Beisel KW, Blake JA, Bradt D, Brusica V, Chothia C, Corbani LE, Cousins S, Dalla E, Dragani TA, Fletcher CF, Forrest A, Frazer KS, Gaasterland T, Gariboldi M, Gissi C, Godzik A, Gough J, Grimmond S, Gustincich S, Hirokawa N, Jackson JJ, Jarvis ED, Kanai A, Kawaji H, Kawasawa Y, Kedzierski RM, King BL, Konagaya A, Kurochkin IV, Lee Y, Lenhard B, Lyons PA, Maglott DR, Maltais L, Marchionni L, McKenzie L, Miki H, Nagashima T, Numata K, Okido T, Pavan WJ, Perteu G, Pesole G, Petrovsky N, Pillai R, Pontius JU, Qi D, Ramachandran S, Ravasi T, Reed JC, Reed DJ, Reid J, Ring BZ, Ringwald M, Sandelin A, Schneider C, Semple CA, Setou M, Shimada K, Sultana R, Takenaka Y, Taylor MS, Teasdale RD, Tomita M, Verardo R, Wagner L, Wahlestedt C, Wang Y, Watanabe Y, Wells C, Wilming LG, Wynshaw-Boris A, Yanagisawa M, Yang I, Yang L, Yuan Z, Zavolan M, Zhu Y, Zimmer A, Carninci P, Hayatsu N, Hirozane-Kishikawa T, Konno H, Nakamura M, Sakazume N, Sato K, Shiraki T, Waki K, Kawai J, Aizawa K, Arakawa T, Fukuda S, Hara A, Hashizume W, Imotani K, Ishii Y, Itoh M, Kagawa I, Miyazaki A, Sakai K, Sasaki D, Shibata K, Shinagawa A, Yasunishi A, Yoshino

- M, Waterston R, Lander ES, Rogers J, Birney E, Hayashizaki Y, Consortium F, I RGERGP, Team II. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*. 2002;420(6915):563-73. Epub 2002/12/06. doi: 10.1038/nature01266. PubMed PMID: 12466851.
52. **Quackenbush J**. Microarray data normalization and transformation. *Nature genetics*. 2002;32 Suppl:496-501. Epub 2002/11/28. doi: 10.1038/ng1032. PubMed PMID: 12454644.
 53. Sonstegard TS, Capuco AV, White J, Van Tassell CP, Connor EE, Cho J, Sultana R, Shade L, Wray JE, Wells KD, **Quackenbush J**. Analysis of bovine mammary gland EST and functional annotation of the *Bos taurus* gene index. *Mammalian genome : official journal of the International Mammalian Genome Society*. 2002;13(7):373-9. Epub 2002/07/26. doi: 10.1007/s00335-001-2145-4. PubMed PMID: 12140684.
 54. Sturn A, **Quackenbush J**, Trajanoski Z. Genesis: cluster analysis of microarray data. *Bioinformatics*. 2002;18(1):207-8. Epub 2002/02/12. PubMed PMID: 11836235.
 55. Yang IV, Chen E, Hasseman JP, Liang W, Frank BC, Wang S, Sharov V, Saeed AI, White J, Li J, Lee NH, Yeatman TJ, **Quackenbush J**. Within the fold: assessing differential expression measures and reproducibility in microarray assays. *Genome biology*. 2002;3(11):research0062. Epub 2002/11/14. PubMed PMID: 12429061; PubMed Central PMCID: PMC133446.
 56. Abulencia JP, Gaspard R, Healy ZR, Gaarde WA, **Quackenbush J**, Konstantopoulos K. Shear-induced cyclooxygenase-2 via a JNK2/c-Jun-dependent pathway regulates prostaglandin receptor expression in chondrocytic cells. *The Journal of biological chemistry*. 2003;278(31):28388-94. Epub 2003/05/14. doi: 10.1074/jbc.M301378200. PubMed PMID: 12743126.
 57. Agrawal D, Chen T, Irby R, **Quackenbush J**, Chambers AF, Szabo M, Cantor A, Coppola D, Yeatman TJ. Osteopontin identified as colon cancer tumor progression marker. *Comptes rendus biologiques*. 2003;326(10-11):1041-3. Epub 2004/01/28. PubMed PMID: 14744111.
 58. Brentani H, Caballero OL, Camargo AA, da Silva AM, da Silva WA, Jr., Dias Neto E, Grivet M, Gruber A, Guimaraes PE, Hide W, Iseli C, Jongeneel CV, Kelso J, Nagai MA, Ojopi EP, Osorio EC, Reis EM, Riggins GJ, Simpson AJ, de Souza S, Stevenson BJ, Strausberg RL, Tajara EH, Verjovski-Almeida S, Acencio ML, Bengtson MH, Bettoni F, Bodmer WF, Briones MR, Camargo LP, Cavenee W, Cerutti JM, Coelho Andrade LE, Costa dos Santos PC, Ramos Costa MC, da Silva IT, Estecio MR, Sa Ferreira K, Furnari FB, Faria M, Jr., Galante PA, Guimaraes GS, Holanda AJ, Kimura ET, Leerkes MR, Lu X, Maciel RM, Martins EA, Massirer KB, Melo AS, Mestriner CA, Miracca EC, Miranda LL, Nobrega FG, Oliveira PS, Paquola AC, Pandolfi JR, Campos Pardini MI, Passetti F, **Quackenbush J**, Schnabel B, Sogayar MC, Souza JE, Valentini SR, Zaiats AC, Amaral EJ, Arnaldi LA, de Araujo AG, de Bessa SA, Bicknell DC, Ribeiro de Camaro ME, Carraro DM, Carrer H, Carvalho AF, Colin C, Costa F, Curcio C, Guerreiro da Silva ID, Pereira da Silva N, Dellamano M, El-Dorry H, Espreafico EM, Scattone Ferreira AJ, Ayres Ferreira C, Fortes MA, Gama AH, Giannella-Neto D, Giannella ML, Giorgi RR, Goldman GH, Goldman MH, Hackel C, Ho PL, Kimura EM, Kowalski LP, Krieger JE, Leite LC, Lopes A, Luna AM, Mackay A, Mari SK, Marques AA, Martins WK, Montagnini A, Mourao Neto M, Nascimento AL, Neville AM, Nobrega MP, O'Hare MJ, Otsuka AY, Ruas de Melo AI, Paco-

- Larson ML, Guimaraes Pereira G, Pereira da Silva N, Pesquero JB, Pessoa JG, Rahal P, Rainho CA, Rodrigues V, Rogatto SR, Romano CM, Romeiro JG, Rossi BM, Rusticci M, Guerra de Sa R, Sant' Anna SC, Sarmazo ML, Silva TC, Soares FA, Sonati Mde F, de Freitas Sousa J, Queiroz D, Valente V, Vettore AL, Villanova FE, Zago MA, Zalcberg H, Human Cancer Genome Project/Cancer Genome Anatomy Project Annotation C, Human Cancer Genome Project Sequencing C. The generation and utilization of a cancer-oriented representation of the human transcriptome by using expressed sequence tags. *Proceedings of the National Academy of Sciences of the United States of America*. 2003;100(23):13418-23. Epub 2003/11/01. doi: 10.1073/pnas.1233632100. PubMed PMID: 14593198; PubMed Central PMCID: PMC263829.
59. Causton HC, **Quackenbush J**, Brazma A. Microarray gene expressions data analysis: a beginner's guide. Malden, MA: Blackwell Pub.; 2003. viii, 160 p.
 60. Chen T, Yang I, Irby R, Shain KH, Wang HG, **Quackenbush J**, Coppola D, Cheng JQ, Yeatman TJ. Regulation of caspase expression and apoptosis by adenomatous polyposis coli. *Cancer research*. 2003;63(15):4368-74. Epub 2003/08/09. PubMed PMID: 12907606.
 61. Cook DN, Wang S, Howles GP, Speer M, Churchhill G, **Quackenbush J**, Schwartz DA. The genetics of innate immunity in the lung. *Chest*. 2003;123(3 Suppl):369S. Epub 2003/03/12. PubMed PMID: 12628980.
 62. Dudoit S, Gentleman RC, **Quackenbush J**. Open source software for the analysis of microarray data. *BioTechniques*. 2003;Suppl:45-51. Epub 2003/04/01. PubMed PMID: 12664684.
 63. Kasukawa T, Furuno M, Nikaido I, Bono H, Hume DA, Bult C, Hill DP, Baldarelli R, Gough J, Kanapin A, Matsuda H, Schriml LM, Hayashizaki Y, Okazaki Y, **Quackenbush J**. Development and evaluation of an automated annotation pipeline and cDNA annotation system. *Genome research*. 2003;13(6B):1542-51. Epub 2003/06/24. doi: 10.1101/gr.992803. PubMed PMID: 12819153; PubMed Central PMCID: PMC403710.
 64. Kim H, Snesrud EC, Haas B, Cheung F, Town CD, **Quackenbush J**. Gene expression analyses of Arabidopsis chromosome 2 using a genomic DNA amplicon microarray. *Genome research*. 2003;13(3):327-40. Epub 2003/03/06. doi: 10.1101/gr.552003. PubMed PMID: 12618363; PubMed Central PMCID: PMC430289.
 65. Klapa MI, **Quackenbush J**. The quest for the mechanisms of life. *Biotechnology and bioengineering*. 2003;84(7):739-42. Epub 2004/01/07. doi: 10.1002/bit.10858. PubMed PMID: 14708113.
 66. Lee Y, **Quackenbush J**. Using the TIGR gene index databases for biological discovery. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]*. 2003;Chapter 1:Unit 1 6. Epub 2008/04/23. doi: 10.1002/0471250953.bi0106s03. PubMed PMID: 18428690.
 67. Merrick JM, Osman A, Tsai J, **Quackenbush J**, LoVerde PT, Lee NH. The *Schistosoma mansoni* gene index: gene discovery and biology by reconstruction and analysis of expressed gene sequences. *The Journal of parasitology*. 2003;89(2):261-9. Epub 2003/05/23. doi: 10.1645/0022-3395(2003)089[0261:TSMGIG]2.0.CO;2. PubMed PMID: 12760639.
 68. Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, **Quackenbush J**. TIGR Gene Indices clustering

- tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*. 2003;19(5):651-2. Epub 2003/03/26. PubMed PMID: 12651724.
69. **Quackenbush J.** John Quackenbush talks about the clinical promise of genetic microarrays. Interviewed by Brian Vastag. *JAMA : the journal of the American Medical Association*. 2003;289(2):159-60, 63. Epub 2003/01/09. PubMed PMID: 12517207.
 70. **Quackenbush J.** Open-source software accelerates bioinformatics. *Genome biology*. 2003;4(9):336. Epub 2003/09/04. doi: 10.1186/gb-2003-4-9-336. PubMed PMID: 12952530; PubMed Central PMCID: PMC193651.
 71. **Quackenbush J.** Genomics. Microarrays--guilt by association. *Science*. 2003;302(5643):240-1. Epub 2003/10/11. doi: 10.1126/science.1090887. PubMed PMID: 14551426.
 72. Rexroad CE, 3rd, Lee Y, Keele JW, Karamycheva S, Brown G, Koop B, Gahr SA, Palti Y, **Quackenbush J.** Sequence analysis of a rainbow trout cDNA library and creation of a gene index. *Cytogenetic and genome research*. 2003;102(1-4):347-54. Epub 2004/02/19. doi: 75773. PubMed PMID: 14970727.
 73. Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J, Pertea GM, Lee Y, Karamycheva S, Sultana R, Tsai J, **Quackenbush J**, Griffiths HM, Restrepo S, Smart CD, Fry WE, Van Der Hoeven R, Tanksley S, Zhang P, Jin H, Yamamoto ML, Baker BJ, Buell CR. Comparative analyses of potato expressed sequence tag libraries. *Plant physiology*. 2003;131(2):419-29. Epub 2003/02/15. doi: 10.1104/pp.013581. PubMed PMID: 12586867; PubMed Central PMCID: PMC166819.
 74. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, **Quackenbush J.** TM4: a free, open-source system for microarray data management and analysis. *BioTechniques*. 2003;34(2):374-8. Epub 2003/03/05. PubMed PMID: 12613259.
 75. Wang HY, Malek RL, Kwitek AE, Greene AS, Luu TV, Behbahani B, Frank B, **Quackenbush J**, Lee NH. Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. *Genome biology*. 2003;4(1):R5. Epub 2003/01/24. PubMed PMID: 12540297; PubMed Central PMCID: PMC151289.
 76. Wells CA, Ravasi T, Sultana R, Yagi K, Carninci P, Bono H, Faulkner G, Okazaki Y, **Quackenbush J**, Hume DA, Lyons PA, Group RG, Members GSL. Continued discovery of transcriptional units expressed in cells of the mouse mononuclear phagocyte lineage. *Genome research*. 2003;13(6B):1360-5. Epub 2003/06/24. doi: 10.1101/gr.1056103. PubMed PMID: 12819134; PubMed Central PMCID: PMC403663.
 77. Whitelaw CA, Barbazuk WB, Pertea G, Chan AP, Cheung F, Lee Y, Zheng L, van Heeringen S, Karamycheva S, Bennetzen JL, SanMiguel P, Lakey N, Bedell J, Yuan Y, Budiman MA, Resnick A, Van Aken S, Utterback T, Riedmuller S, Williams M, Feldblyum T, Schubert K, Beachy R, Fraser CM, **Quackenbush J.** Enrichment of gene-coding sequences in maize by genome filtration. *Science*. 2003;302(5653):2118-20. Epub 2003/12/20. doi: 10.1126/science.1090047. PubMed PMID: 14684821.
 78. Yuan Q, Ouyang S, Liu J, Suh B, Cheung F, Sultana R, Lee D, **Quackenbush J**, Buell CR. The TIGR rice genome annotation resource: annotating the rice genome and creating resources for plant biologists. *Nucleic acids research*. 2003;31(1):229-

33. Epub 2003/01/10. PubMed PMID: 12519988; PubMed Central PMCID: PMC165506.
79. Zhu Y, King BL, Parvizi B, Brunk BP, Stoeckert CJ, Jr., **Quackenbush J**, Richardson J, Bult CJ. Integrating computationally assembled mouse transcript sequences with the Mouse Genome Informatics (MGI) database. *Genome biology*. 2003;4(2):R16. Epub 2003/03/07. PubMed PMID: 12620126; PubMed Central PMCID: PMC151306.
80. Ball C, Brazma A, Causton H, Chervitz S, Edgar R, Hingamp P, Matese JC, Icahn C, Parkinson H, **Quackenbush J**, Ringwald M, Sansone SA, Sherlock G, Spellman P, Stoeckert C, Tatenio Y, Taylor R, White J, Winegarden N, Microarray Gene Expression Data S. An open letter on microarray data from the MGED Society. *Microbiology*. 2004;150(Pt 11):3522-4. Epub 2004/11/06. doi: 10.1099/mic.0.27637-0. PubMed PMID: 15528642.
81. Ball C, Brazma A, Causton H, Chervitz S, Edgar R, Hingamp P, Matese JC, Parkinson H, **Quackenbush J**, Ringwald M, Sansone SA, Sherlock G, Spellman P, Stoeckert C, Tatenio Y, Taylor R, White J, Winegarden N, Society M. Standards for microarray data: an open letter. *Environmental health perspectives*. 2004;112(12):A666-7. Epub 2004/09/04. PubMed PMID: 15345376; PubMed Central PMCID: PMC1277123.
82. Ball CA, Brazma A, Causton H, Chervitz S, Edgar R, Hingamp P, Matese JC, Parkinson H, **Quackenbush J**, Ringwald M, Sansone SA, Sherlock G, Spellman P, Stoeckert C, Tatenio Y, Taylor R, White J, Winegarden N. Submission of microarray data to public repositories. *PLoS biology*. 2004;2(9):E317. Epub 2004/09/02. doi: 10.1371/journal.pbio.0020317. PubMed PMID: 15340489; PubMed Central PMCID: PMC514887.
83. Bloom G, Yang IV, Boulware D, Kwong KY, Coppola D, Eschrich S, **Quackenbush J**, Yeatman TJ. Multi-platform, multi-site, microarray-based human tumor classification. *The American journal of pathology*. 2004;164(1):9-16. Epub 2003/12/26. doi: 10.1016/S0002-9440(10)63090-8. PubMed PMID: 14695313; PubMed Central PMCID: PMC1602228.
84. Cook DN, Wang S, Wang Y, Howles GP, Whitehead GS, Berman KG, Church TD, Frank BC, Gaspard RM, Yu Y, **Quackenbush J**, Schwartz DA. Genetic regulation of endotoxin-induced airway disease. *Genomics*. 2004;83(6):961-9. Epub 2004/06/05. doi: 10.1016/j.ygeno.2003.12.008. PubMed PMID: 15177550.
85. Cronin M, Ghosh K, Sistare F, **Quackenbush J**, Vilker V, O'Connell C. Universal RNA reference materials for gene expression. *Clinical chemistry*. 2004;50(8):1464-71. Epub 2004/05/25. doi: 10.1373/clinchem.2004.035675. PubMed PMID: 15155546.
86. Ghedin E, Pumfery A, de la Fuente C, Yao K, Miller N, Lacoste V, **Quackenbush J**, Jacobson S, Kashanchi F. Use of a multi-virus array for the study of human viral and retroviral pathogens: gene expression studies and ChIP-chip analysis. *Retrovirology*. 2004;1:10. Epub 2004/06/01. doi: 10.1186/1742-4690-1-10. PubMed PMID: 15169557; PubMed Central PMCID: PMC442135.
87. Grigoryev DN, Ma SF, Irizarry RA, Ye SQ, **Quackenbush J**, Garcia JG. Orthologous gene-expression profiling in multi-species models: search for candidate genes. *Genome biology*. 2004;5(5):R34. Epub 2004/05/07. doi: 10.1186/gb-2004-5-5-r34. PubMed PMID: 15128448; PubMed Central PMCID: PMC416470.

88. Imanishi T, Itoh T, Suzuki Y, O'Donovan C, Fukuchi S, Koyanagi KO, Barrero RA, Tamura T, Yamaguchi-Kabata Y, Tanino M, Yura K, Miyazaki S, Ikeo K, Homma K, Kasprzyk A, Nishikawa T, Hirakawa M, Thierry-Mieg J, Thierry-Mieg D, Ashurst J, Jia L, Nakao M, Thomas MA, Mulder N, Karavidopoulou Y, Jin L, Kim S, Yasuda T, Lenhard B, Eveno E, Suzuki Y, Yamasaki C, Takeda J, Gough C, Hilton P, Fujii Y, Sakai H, Tanaka S, Amid C, Bellgard M, Bonaldo Mde F, Bono H, Bromberg SK, Brookes AJ, Bruford E, Carninci P, Chelala C, Couillault C, de Souza SJ, Debily MA, Devignes MD, Dubchak I, Endo T, Estreicher A, Eyras E, Fukami-Kobayashi K, Gopinath GR, Graudens E, Hahn Y, Han M, Han ZG, Hanada K, Hanaoka H, Harada E, Hashimoto K, Hinz U, Hirai M, Hishiki T, Hopkinson I, Imbeaud S, Inoko H, Kanapin A, Kaneko Y, Kasukawa T, Kelso J, Kersey P, Kikuno R, Kimura K, Korn B, Kuryshv V, Makalowska I, Makino T, Mano S, Mariage-Samson R, Mashima J, Matsuda H, Mewes HW, Minoshima S, Nagai K, Nagasaki H, Nagata N, Nigam R, Ogasawara O, Ohara O, Ohtsubo M, Okada N, Okido T, Oota S, Ota M, Ota T, Otsuki T, Piatier-Tonneau D, Poustka A, Ren SX, Saitou N, Sakai K, Sakamoto S, Sakate R, Schupp I, Servant F, Sherry S, Shiba R, Shimizu N, Shimoyama M, Simpson AJ, Soares B, Steward C, Suwa M, Suzuki M, Takahashi A, Tamiya G, Tanaka H, Taylor T, Terwilliger JD, Unneberg P, Veeramachaneni V, Watanabe S, Wilming L, Yasuda N, Yoo HS, Stodolsky M, Makalowski W, Go M, Nakai K, Takagi T, Kanehisa M, Sakaki Y, **Quackenbush J**, Okazaki Y, Hayashizaki Y, Hide W, Chakraborty R, Nishikawa K, Sugawara H, Tateno Y, Chen Z, Oishi M, Tonellato P, Apweiler R, Okubo K, Wagner L, Wiemann S, Strausberg RL, Isogai T, Auffray C, Nomura N, Gojobori T, Sugano S. Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS biology*. 2004;2(6):e162. Epub 2004/04/23. doi: 10.1371/journal.pbio.0020162. PubMed PMID: 15103394; PubMed Central PMCID: PMC393292.
89. Larkin JE, Frank BC, Gaspard RM, Duka I, Gavras H, **Quackenbush J**. Cardiac transcriptional response to acute and chronic angiotensin II treatments. *Physiological genomics*. 2004;18(2):152-66. Epub 2004/05/06. doi: 10.1152/physiolgenomics.00057.2004. PubMed PMID: 15126644.
90. Nene V, Lee D, Kang'a S, Skilton R, Shah T, de Villiers E, Mwaura S, Taylor D, **Quackenbush J**, Bishop R. Genes transcribed in the salivary glands of female *Rhipicephalus appendiculatus* ticks infected with *Theileria parva*. *Insect biochemistry and molecular biology*. 2004;34(10):1117-28. Epub 2004/10/12. doi: 10.1016/j.ibmb.2004.07.002. PubMed PMID: 15475305.
91. **Quackenbush J**. Data standards for 'omic' science. *Nature biotechnology*. 2004;22(5):613-4. Epub 2004/05/04. doi: 10.1038/nbt0504-613. PubMed PMID: 15122299.
92. Sharov V, Kwong KY, Frank B, Chen E, Hasseman J, Gaspard R, Yu Y, Yang I, **Quackenbush J**. The limits of log-ratios. *BMC biotechnology*. 2004;4:3. Epub 2004/04/29. doi: 10.1186/1472-6750-4-3. PubMed PMID: 15113428; PubMed Central PMCID: PMC400743.
93. Bergman NH, Passalacqua KD, Gaspard R, Shetron-Rama LM, **Quackenbush J**, Hanna PC. Murine macrophage transcriptional responses to *Bacillus anthracis* infection and intoxication. *Infection and immunity*. 2005;73(2):1069-80. Epub 2005/01/25. doi: 10.1128/IAI.73.2.1069-1080.2005. PubMed PMID: 15664951; PubMed Central PMCID: PMC547061.

94. 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-Impimbato 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, Brusica 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, Consortium F, Group RGER, Genome Science G. The transcriptional landscape of the mammalian genome. *Science*. 2005;309(5740):1559-63. Epub 2005/09/06. doi: 10.1126/science.1112014. PubMed PMID: 16141072.
95. Cui L, Fan Q, Hu Y, Karamycheva SA, **Quackenbush J**, Khuntirat B, Sattabongkot J, Carlton JM. Gene discovery in *Plasmodium vivax* through sequencing of ESTs from mixed blood stages. *Molecular and biochemical parasitology*. 2005;144(1):1-9. Epub 2005/08/09. doi: 10.1016/j.molbiopara.2005.05.016. PubMed PMID: 16085323.
96. Djebbari A, Karamycheva S, Howe E, **Quackenbush J**. MeSHer: identifying biological concepts in microarray assays based on PubMed references and MeSH terms. *Bioinformatics*. 2005;21(15):3324-6. Epub 2005/05/28. doi: 10.1093/bioinformatics/bti503. PubMed PMID: 15919728.
97. Eschrich S, Yang I, Bloom G, Kwong KY, Boulware D, Cantor A, Coppola D, Kruhboffer M, Aaltonen L, Orntoft TF, **Quackenbush J**, Yeatman TJ. Molecular staging for survival prediction of colorectal cancer patients. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2005;23(15):3526-35. Epub 2005/05/24. doi: 10.1200/JCO.2005.00.695. PubMed PMID: 15908663.

98. Guerrero FD, Miller RJ, Rousseau ME, Sunkara S, **Quackenbush J**, Lee Y, Nene V. BmiGI: a database of cDNAs expressed in *Boophilus microplus*, the tropical/southern cattle tick. *Insect biochemistry and molecular biology*. 2005;35(6):585-95. Epub 2005/04/29. doi: 10.1016/j.ibmb.2005.01.020. PubMed PMID: 15857764.
99. Hackl H, Burkard TR, Sturn A, Rubio R, Schleiffer A, Tian S, **Quackenbush J**, Eisenhaber F, Trajanoski Z. Molecular processes during fat cell development revealed by gene expression profiling and functional annotation. *Genome biology*. 2005;6(13):R108. Epub 2006/01/20. doi: 10.1186/gb-2005-6-13-r108. PubMed PMID: 16420668; PubMed Central PMCID: PMC1414107.
100. Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J, Germino G, Griffin C, Hilmer SC, Hoffman E, Jedlicka AE, Kawasaki E, Martinez-Murillo F, Morsberger L, Lee H, Petersen D, **Quackenbush J**, Scott A, Wilson M, Yang Y, Ye SQ, Yu W. Multiple-laboratory comparison of microarray platforms. *Nature methods*. 2005;2(5):345-50. Epub 2005/04/23. doi: 10.1038/nmeth756. PubMed PMID: 15846361.
101. Kwong KY, Bloom GC, Yang I, Boulware D, Coppola D, Haseman J, Chen E, McGrath A, Makusky AJ, Taylor J, Steiner S, Zhou J, Yeatman TJ, **Quackenbush J**. Synchronous global assessment of gene and protein expression in colorectal cancer progression. *Genomics*. 2005;86(2):142-58. Epub 2005/06/14. doi: 10.1016/j.ygeno.2005.03.012. PubMed PMID: 15951154.
102. Larkin JE, Frank BC, Gavras H, Sultana R, **Quackenbush J**. Independence and reproducibility across microarray platforms. *Nature methods*. 2005;2(5):337-44. Epub 2005/04/23. doi: 10.1038/nmeth757. PubMed PMID: 15846360.
103. Lee Y, Tsai J, Sunkara S, Karamycheva S, Pertea G, Sultana R, Antonescu V, Chan A, Cheung F, **Quackenbush J**. The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic acids research*. 2005;33(Database issue):D71-4. Epub 2004/12/21. doi: 10.1093/nar/gki064. PubMed PMID: 15608288; PubMed Central PMCID: PMC540018.
104. Liu F, Vantoai T, Moy LP, Bock G, Linford LD, **Quackenbush J**. Global transcription profiling reveals comprehensive insights into hypoxic response in *Arabidopsis*. *Plant physiology*. 2005;137(3):1115-29. Epub 2005/03/01. doi: 10.1104/pp.104.055475. PubMed PMID: 15734912; PubMed Central PMCID: PMC1065411.
105. Margolin AA, Greshock J, Naylor TL, Mosse Y, Maris JM, Bignell G, Saeed AI, **Quackenbush J**, Weber BL. CGHAnalyzer: a stand-alone software package for cancer genome analysis using array-based DNA copy number data. *Bioinformatics*. 2005;21(15):3308-11. Epub 2005/05/21. doi: 10.1093/bioinformatics/bti500. PubMed PMID: 15905276.
106. Marko NF, Frank B, **Quackenbush J**, Lee NH. A robust method for the amplification of RNA in the sense orientation. *BMC genomics*. 2005;6:27. Epub 2005/03/03. doi: 10.1186/1471-2164-6-27. PubMed PMID: 15740627; PubMed Central PMCID: PMC554769.
107. Pinet M, Blade MC, Salvado MJ, Arola L, Hackl H, **Quackenbush J**, Trajanoski Z, Ardevol A. Grape-seed derived procyanidins interfere with adipogenesis of 3T3-L1 cells at the onset of differentiation. *Int J Obes (Lond)*. 2005;29(8):934-41. Epub 2005/05/27. doi: 10.1038/sj.ijo.0802988. PubMed PMID: 15917849.

108. **Quackenbush J.** Extracting meaning from functional genomics experiments. *Toxicology and applied pharmacology*. 2005;207(2 Suppl):195-9. Epub 2005/07/09. doi: 10.1016/j.taap.2005.04.029. PubMed PMID: 16002114.
109. Stanley Kim H, Yu Y, Snedrud EC, Moy LP, Linford LD, Haas BJ, Nierman WC, **Quackenbush J.** Transcriptional divergence of the duplicated oxidative stress-responsive genes in the Arabidopsis genome. *The Plant journal : for cell and molecular biology*. 2005;41(2):212-20. Epub 2005/01/07. doi: 10.1111/j.1365-3113X.2004.02295.x. PubMed PMID: 15634198.
110. Brooksbank C, **Quackenbush J.** Data standards: a call to action. *Omics : a journal of integrative biology*. 2006;10(2):94-9. Epub 2006/08/12. doi: 10.1089/omi.2006.10.94. PubMed PMID: 16901212.
111. Bruggmann R, Bharti AK, Gundlach H, Lai J, Young S, Pontaroli AC, Wei F, Haberer G, Fuks G, Du C, Raymond C, Estep MC, Liu R, Bennetzen JL, Chan AP, Rabinowicz PD, **Quackenbush J**, Barbazuk WB, Wing RA, Birren B, Nusbaum C, Rounsley S, Mayer KF, Messing J. Uneven chromosome contraction and expansion in the maize genome. *Genome research*. 2006;16(10):1241-51. Epub 2006/08/12. doi: 10.1101/gr.5338906. PubMed PMID: 16902087; PubMed Central PMCID: PMC1581433.
112. Chan AP, Pertea G, Cheung F, Lee D, Zheng L, Whitelaw C, Pontaroli AC, SanMiguel P, Yuan Y, Bennetzen J, Barbazuk WB, **Quackenbush J**, Rabinowicz PD. The TIGR Maize Database. *Nucleic acids research*. 2006;34(Database issue):D771-6. Epub 2005/12/31. doi: 10.1093/nar/gkj072. PubMed PMID: 16381977; PubMed Central PMCID: PMC1347435.
113. Chan VS, Chau SY, Tian L, Chen Y, Kwong SK, **Quackenbush J**, Dallman M, Lamb J, Tam PK. Sonic hedgehog promotes CD4⁺ T lymphocyte proliferation and modulates the expression of a subset of CD28-targeted genes. *International immunology*. 2006;18(12):1627-36. Epub 2006/09/29. doi: 10.1093/intimm/dxl096. PubMed PMID: 17005629.
114. Deutsch EW, Ball CA, Bova GS, Brazma A, Bumgarner RE, Campbell D, Causton HC, Christiansen J, Davidson D, Eichner LJ, Goo YA, Grimmond S, Henrich T, Johnson MH, Korb M, Mills JC, Oudes A, Parkinson HE, Pascal LE, **Quackenbush J**, Ramialison M, Ringwald M, Sansone SA, Sherlock G, Stoeckert CJ, Jr., Swedlow J, Taylor RC, Walashek L, Zhou Y, Liu AY, True LD. Development of the Minimum Information Specification for In Situ Hybridization and Immunohistochemistry Experiments (MISFISHIE). *Omics : a journal of integrative biology*. 2006;10(2):205-8. Epub 2006/08/12. doi: 10.1089/omi.2006.10.205. PubMed PMID: 16901227.
115. Frith MC, Bailey TL, Kasukawa T, Mignone F, Kummerfeld SK, Madera M, Sunkara S, Furuno M, Bult CJ, **Quackenbush J**, Kai C, Kawai J, Carninci P, Hayashizaki Y, Pesole G, Mattick JS. Discrimination of non-protein-coding transcripts from protein-coding mRNA. *RNA biology*. 2006;3(1):40-8. Epub 2006/11/23. PubMed PMID: 17114936.
116. Hu VW, Frank BC, Heine S, Lee NH, **Quackenbush J.** Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. *BMC genomics*. 2006;7:118. Epub 2006/05/20. doi: 10.1186/1471-2164-7-118. PubMed PMID: 16709250; PubMed Central PMCID: PMC1525191.

117. Maeda N, Kasukawa T, Oyama R, Gough J, Frith M, Engstrom PG, Lenhard B, Aturaliya RN, Batalov S, Beisel KW, Bult CJ, Fletcher CF, Forrest AR, Furuno M, Hill D, Itoh M, Kanamori-Katayama M, Katayama S, Katoh M, Kawashima T, **Quackenbush J**, Ravasi T, Ring BZ, Shibata K, Sugiura K, Takenaka Y, Teasdale RD, Wells CA, Zhu Y, Kai C, Kawai J, Hume DA, Carninci P, Hayashizaki Y. Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. *PLoS genetics*. 2006;2(4):e62. Epub 2006/05/10. doi: 10.1371/journal.pgen.0020062. PubMed PMID: 16683036; PubMed Central PMCID: PMC1449903.
118. Maher EA, Brennan C, Wen PY, Durso L, Ligon KL, Richardson A, Khatry D, Feng B, Sinha R, Louis DN, **Quackenbush J**, Black PM, Chin L, DePinho RA. Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. *Cancer research*. 2006;66(23):11502-13. Epub 2006/11/23. doi: 10.1158/0008-5472.CAN-06-2072. PubMed PMID: 17114236.
119. Malek RL, Wang HY, Kwitek AE, Greene AS, Bhagabati N, Borchardt G, Cahill L, Currier T, Frank B, Fu X, Hasinoff M, Howe E, Letwin N, Luu TV, Saeed A, Sajadi H, Salzberg SL, Sultana R, Thiagarajan M, Tsai J, Veratti K, White J, **Quackenbush J**, Jacob HJ, Lee NH. Physiogenomic resources for rat models of heart, lung and blood disorders. *Nature genetics*. 2006;38(2):234-9. Epub 2006/01/18. doi: 10.1038/ng1693. PubMed PMID: 16415889.
120. Mar JC, Rubio R, **Quackenbush J**. Inferring steady state single-cell gene expression distributions from analysis of mesoscopic samples. *Genome biology*. 2006;7(12):R119. Epub 2006/12/16. doi: 10.1186/gb-2006-7-12-r119. PubMed PMID: 17169148; PubMed Central PMCID: PMC1794432.
121. **Quackenbush J**. Standardizing the standards. *Molecular systems biology*. 2006;2:2006 0010. Epub 2006/06/02. doi: 10.1038/msb4100052. PubMed PMID: 16738556; PubMed Central PMCID: PMC1681484.
122. **Quackenbush J**. Microarray analysis and tumor classification. *The New England journal of medicine*. 2006;354(23):2463-72. Epub 2006/06/09. doi: 10.1056/NEJMra042342. PubMed PMID: 16760446.
123. **Quackenbush J**. From 'omes to biology. *Animal genetics*. 2006;37 Suppl 1:48-56. Epub 2006/08/05. doi: 10.1111/j.1365-2052.2006.01476.x. PubMed PMID: 16887002.
124. **Quackenbush J**. Computational approaches to analysis of DNA microarray data. *Yearbook of medical informatics*. 2006:91-103. Epub 2006/10/20. PubMed PMID: 17051302.
125. **Quackenbush J**. Weighing our measures of gene expression. *Molecular systems biology*. 2006;2:63. Epub 2006/11/15. doi: 10.1038/msb4100096. PubMed PMID: 17102808; PubMed Central PMCID: PMC1682020.
126. **Quackenbush J**, Irizarry RA. Response to Shields: 'MIAME, we have a problem'. *Trends in genetics : TIG*. 2006;22(9):471-2. Epub 2006/07/25. doi: 10.1016/j.tig.2006.07.007. PubMed PMID: 16859802.
127. **Quackenbush J**, Salzberg SL. It is time to end the patenting of software. *Bioinformatics*. 2006;22(12):1416-7. Epub 2006/06/13. doi: 10.1093/bioinformatics/btl167. PubMed PMID: 16766564; PubMed Central PMCID: PMC2836512.

128. **Quackenbush J**, Stoeckert C, Ball C, Brazma A, Gentleman R, Huber W, Irizarry R, Salit M, Sherlock G, Spellman P, Winegarden N. Top-down standards will not serve systems biology. *Nature*. 2006;440(7080):24. Epub 2006/03/03. doi: 10.1038/440024a. PubMed PMID: 16511469.
129. Rayner TF, Rocca-Serra P, Spellman PT, Causton HC, Farne A, Holloway E, Irizarry RA, Liu J, Maier DS, Miller M, Petersen K, **Quackenbush J**, Sherlock G, Stoeckert CJ, Jr., White J, Whetzel PL, Wymore F, Parkinson H, Sarkans U, Ball CA, Brazma A. A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. *BMC bioinformatics*. 2006;7:489. Epub 2006/11/08. doi: 10.1186/1471-2105-7-489. PubMed PMID: 17087822; PubMed Central PMCID: PMC1687205.
130. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, **Quackenbush J**. TM4 microarray software suite. *Methods in enzymology*. 2006;411:134-93. Epub 2006/08/31. doi: 10.1016/S0076-6879(06)11009-5. PubMed PMID: 16939790.
131. Stoeckert C, Ball C, Brazma A, Brinkman R, Causton H, Fan L, Fostel J, Fragoso G, Heiskanen M, Holstege F, Morrison N, Parkinson H, **Quackenbush J**, Rocca-Serra P, Sansone SA, Sarkans U, Sherlock G, Stevens R, Taylor C, Taylor R, Whetzel P, White J. Wrestling with SUMO and bio-ontologies. *Nature biotechnology*. 2006;24(1):21-2; author reply 3. Epub 2006/01/13. doi: 10.1038/nbt0106-21a. PubMed PMID: 16404382.
132. Walker JK, Ahumada A, Frank B, Gaspard R, Berman K, **Quackenbush J**, Schwartz DA. Multistrain genetic comparisons reveal CCR5 as a receptor involved in airway hyperresponsiveness. *American journal of respiratory cell and molecular biology*. 2006;34(6):711-8. Epub 2006/02/14. doi: 10.1165/rcmb.2005-0314OC. PubMed PMID: 16474097; PubMed Central PMCID: PMC2644233.
133. Chan AP, Rabinowicz PD, **Quackenbush J**, Buell CR, Town CD. Plant database resources at The Institute for Genomic Research. *Methods Mol Biol*. 2007;406:113-36. Epub 2008/02/22. PubMed PMID: 18287690.
134. Danley PD, Mullen SP, Liu F, Nene V, **Quackenbush J**, Shaw KL. A cricket Gene Index: a genomic resource for studying neurobiology, speciation, and molecular evolution. *BMC genomics*. 2007;8:109. Epub 2007/04/27. doi: 10.1186/1471-2164-8-109. PubMed PMID: 17459168; PubMed Central PMCID: PMC1878485.
135. Hollingsworth JW, Whitehead G, Berman KG, Tekippe EM, Gilmour MI, Larkin JE, **Quackenbush J**, Schwartz DA. Genetic basis of murine antibacterial defense to streptococcal lung infection. *Immunogenetics*. 2007;59(9):713-24. Epub 2007/08/19. doi: 10.1007/s00251-007-0242-6. PubMed PMID: 17701033.
136. **Quackenbush J**. Extracting biology from high-dimensional biological data. *The Journal of experimental biology*. 2007;210(Pt 9):1507-17. Epub 2007/04/24. doi: 10.1242/jeb.004432. PubMed PMID: 17449816.
137. Raj JU, Aliferis C, Caprioli RM, Cowley AW, Jr., Davies PF, Duncan MW, Erle DJ, Erzurum SC, Finn PW, Ischiropoulos H, Kaminski N, Kleeberger SR, Leikauf GD, Loyd JE, Martin TR, Matalon S, Moore JH, **Quackenbush J**, Sabo-Attwood T, Shapiro SD, Schnitzer JE, Schwartz DA, Schwiebert LM, Sheppard D, Ware LB, Weiss ST, Whitsett JA, Wurfel MM, Matthay MA. Genomics and proteomics of lung disease: conference summary. *American journal of physiology Lung cellular and molecular physiology*. 2007;293(1):L45-51. Epub 2007/05/01. doi:

- 10.1152/ajplung.00139.2007. PubMed PMID: 17468134; PubMed Central PMCID: PMC4212816.
138. Song JS, Maghsoudi K, Li W, Fox E, **Quackenbush J**, Shirley Liu X. Microarray blob-defect removal improves array analysis. *Bioinformatics*. 2007;23(8):966-71. Epub 2007/03/03. doi: 10.1093/bioinformatics/btm043. PubMed PMID: 17332024.
 139. Aryee MJ, **Quackenbush J**. An optimized predictive strategy for interactome mapping. *Journal of proteome research*. 2008;7(9):4089-94. Epub 2008/07/23. doi: 10.1021/pr700858e. PubMed PMID: 18642945.
 140. Boedigheimer MJ, Wolfinger RD, Bass MB, Bushel PR, Chou JW, Cooper M, Corton JC, Fostel J, Hester S, Lee JS, Liu F, Liu J, Qian HR, **Quackenbush J**, Pettit S, Thompson KL. Sources of variation in baseline gene expression levels from toxicogenomics study control animals across multiple laboratories. *BMC genomics*. 2008;9:285. Epub 2008/06/14. doi: 10.1186/1471-2164-9-285. PubMed PMID: 18549499; PubMed Central PMCID: PMC2453529.
 141. Cheng AS, Culhane AC, Chan MW, Venkataramu CR, Ehrich M, Nasir A, Rodriguez BA, Liu J, Yan PS, **Quackenbush J**, Nephew KP, Yeatman TJ, Huang TH. Epithelial progeny of estrogen-exposed breast progenitor cells display a cancer-like methylome. *Cancer research*. 2008;68(6):1786-96. Epub 2008/03/15. doi: 10.1158/0008-5472.CAN-07-5547. PubMed PMID: 18339859; PubMed Central PMCID: PMC4172329.
 142. Chittenden TW, Howe EA, Culhane AC, Sultana R, Taylor JM, Holmes C, **Quackenbush J**. Functional classification analysis of somatically mutated genes in human breast and colorectal cancers. *Genomics*. 2008;91(6):508-11. Epub 2008/04/25. doi: 10.1016/j.ygeno.2008.03.002. PubMed PMID: 18434084; PubMed Central PMCID: PMC2492759.
 143. Deutsch EW, Ball CA, Berman JJ, Bova GS, Brazma A, Bumgarner RE, Campbell D, Causton HC, Christiansen JH, Daian F, Dauga D, Davidson DR, Gimenez G, Goo YA, Grimmond S, Henrich T, Herrmann BG, Johnson MH, Korb M, Mills JC, Oudes AJ, Parkinson HE, Pascal LE, Pollet N, **Quackenbush J**, Ramialison M, Ringwald M, Salgado D, Sansone SA, Sherlock G, Stoeckert CJ, Jr., Swedlow J, Taylor RC, Walashek L, Warford A, Wilkinson DG, Zhou Y, Zon LI, Liu AY, True LD. Minimum information specification for in situ hybridization and immunohistochemistry experiments (MISFISHIE). *Nature biotechnology*. 2008;26(3):305-12. Epub 2008/03/11. doi: 10.1038/nbt1391. PubMed PMID: 18327244; PubMed Central PMCID: PMC4367930.
 144. Djebbari A, **Quackenbush J**. Seeded Bayesian Networks: constructing genetic networks from microarray data. *BMC systems biology*. 2008;2:57. Epub 2008/07/08. doi: 10.1186/1752-0509-2-57. PubMed PMID: 18601736; PubMed Central PMCID: PMC2474592.
 145. Kahvejian A, **Quackenbush J**, Thompson JF. What would you do if you could sequence everything? *Nature biotechnology*. 2008;26(10):1125-33. Epub 2008/10/11. doi: 10.1038/nbt1494. PubMed PMID: 18846086; PubMed Central PMCID: PMC4153598.
 146. Kitami T, Rubio R, O'Brien W, **Quackenbush J**, Nadeau JH. Gene-environment interactions reveal a homeostatic role for cholesterol metabolism during dietary folate perturbation in mice. *Physiological genomics*. 2008;35(2):182-90. Epub 2008/08/14. doi: 10.1152/physiolgenomics.00294.2007. PubMed PMID: 18697859; PubMed Central PMCID: PMC2575958.

147. Nosho K, Irahara N, Shima K, Kure S, Kirkner GJ, Schernhammer ES, Hazra A, Hunter DJ, **Quackenbush J**, Spiegelman D, Giovannucci EL, Fuchs CS, Ogino S. Comprehensive biostatistical analysis of CpG island methylator phenotype in colorectal cancer using a large population-based sample. *PloS one*. 2008;3(11):e3698. Epub 2008/11/13. doi: 10.1371/journal.pone.0003698. PubMed PMID: 19002263; PubMed Central PMCID: PMC2579485.
148. Taoufik E, Petit E, Divoux D, Tseveleki V, Mengozzi M, Roberts ML, Valable S, Ghezzi P, **Quackenbush J**, Brines M, Cerami A, Probert L. TNF receptor I sensitizes neurons to erythropoietin- and VEGF-mediated neuroprotection after ischemic and excitotoxic injury. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(16):6185-90. Epub 2008/04/17. doi: 10.1073/pnas.0801447105. PubMed PMID: 18413601; PubMed Central PMCID: PMC2299225.
149. Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz PA, Bogue M, Booth T, Brazma A, Brinkman RR, Michael Clark A, Deutsch EW, Fiehn O, Fostel J, Ghazal P, Gibson F, Gray T, Grimes G, Hancock JM, Hardy NW, Hermjakob H, Julian RK, Jr., Kane M, Kettner C, Kinsinger C, Kolker E, Kuiper M, Le Novere N, Leebens-Mack J, Lewis SE, Lord P, Mallon AM, Marthandan N, Masuya H, McNally R, Mehrle A, Morrison N, Orchard S, **Quackenbush J**, Reecy JM, Robertson DG, Rocca-Serra P, Rodriguez H, Rosenfelder H, Santoyo-Lopez J, Scheuermann RH, Schober D, Smith B, Snape J, Stoeckert CJ, Jr., Tipton K, Sterk P, Untergasser A, Vandesompele J, Wiemann S. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nature biotechnology*. 2008;26(8):889-96. Epub 2008/08/09. doi: 10.1038/nbt.1411. PubMed PMID: 18688244; PubMed Central PMCID: PMC2771753.
150. April C, Klotzle B, Royce T, Wickham-Garcia E, Boyaniwsky T, Izzo J, Cox D, Jones W, Rubio R, Holton K, Matulonis U, **Quackenbush J**, Fan JB. Whole-genome gene expression profiling of formalin-fixed, paraffin-embedded tissue samples. *PloS one*. 2009;4(12):e8162. Epub 2009/12/10. doi: 10.1371/journal.pone.0008162. PubMed PMID: 19997620; PubMed Central PMCID: PMC2780295.
151. Aryee MJ, Gutierrez-Pabello JA, Kramnik I, Maiti T, **Quackenbush J**. An improved empirical bayes approach to estimating differential gene expression in microarray time-course data: BETR (Bayesian Estimation of Temporal Regulation). *BMC bioinformatics*. 2009;10:409. Epub 2009/12/17. doi: 10.1186/1471-2105-10-409. PubMed PMID: 20003283; PubMed Central PMCID: PMC2801687.
152. Bateman A, **Quackenbush J**. Bioinformatics for next generation sequencing. *Bioinformatics*. 2009;25(4):429. Epub 2009/02/10. doi: 10.1093/bioinformatics/btp037. PubMed PMID: 19202193.
153. Bodingheimer MJ, Chou JW, Corton JC, Fostel J, O'Lone R, Pine PS, **Quackenbush J**, Thompson KL, Wolfinger RD. Variance due to Smooth Bias in Rat Liver and Kidney Baseline Gene Expression in a Large Multi-laboratory Data Set. In: Scherer A, editor. *Batch Effects and Noise in Microarray Experiments: Sources and Solutions* John Wiley and Sons, Ltd; 2009. p. 87-99.
154. Colak D, Kaya N, Al-Zahrani J, Al Bakheet A, Muiya P, Andres E, **Quackenbush J**, Dzimiri N. Left ventricular global transcriptional profiling in human end-stage dilated cardiomyopathy. *Genomics*. 2009;94(1):20-31. Epub 2009/04/01. doi:

- 10.1016/j.ygeno.2009.03.003. PubMed PMID: 19332114; PubMed Central PMCID: PMC4152850.
155. Consortium F, Suzuki H, Forrest AR, van Nimwegen E, Daub CO, Balwiercz PJ, Irvine KM, Lassmann T, Ravasi T, Hasegawa Y, de Hoon MJ, Katayama S, Schroder K, Carninci P, Tomaru Y, Kanamori-Katayama M, Kubosaki A, Akalin A, Ando Y, Arner E, Asada M, Asahara H, Bailey T, Bajic VB, Bauer D, Beckhouse AG, Bertin N, Bjorkegren J, Brombacher F, Bulger E, Chalk AM, Chiba J, Cloonan N, Dawe A, Dostie J, Engstrom PG, Essack M, Faulkner GJ, Fink JL, Fredman D, Fujimori K, Furuno M, Gojobori T, Gough J, Grimmond SM, Gustafsson M, Hashimoto M, Hashimoto T, Hatakeyama M, Heinzl S, Hide W, Hofmann O, Hornquist M, Huminiecki L, Ikeo K, Imamoto N, Inoue S, Inoue Y, Ishihara R, Iwayanagi T, Jacobsen A, Kaur M, Kawaji H, Kerr MC, Kimura R, Kimura S, Kimura Y, Kitano H, Koga H, Kojima T, Kondo S, Konno T, Krogh A, Kruger A, Kumar A, Lenhard B, Lennartsson A, Lindow M, Lizio M, Macpherson C, Maeda N, Maher CA, Maqungo M, Mar J, Matigian NA, Matsuda H, Mattick JS, Meier S, Miyamoto S, Miyamoto-Sato E, Nakabayashi K, Nakachi Y, Nakano M, Nygaard S, Okayama T, Okazaki Y, Okuda-Yabukami H, Orlando V, Otomo J, Pachkov M, Petrovsky N, Plessy C, **Quackenbush J**, Radovanovic A, Rehli M, Saito R, Sandelin A, Schmeier S, Schonbach C, Schwartz AS, Semple CA, Sera M, Severin J, Shirahige K, Simons C, St Laurent G, Suzuki M, Suzuki T, Sweet MJ, Taft RJ, Takeda S, Takenaka Y, Tan K, Taylor MS, Teasdale RD, Tegner J, Teichmann S, Valen E, Wahlestedt C, Waki K, Waterhouse A, Wells CA, Winther O, Wu L, Yamaguchi K, Yanagawa H, Yasuda J, Zavolan M, Hume DA, Riken Omics Science C, Arakawa T, Fukuda S, Imamura K, Kai C, Kaiho A, Kawashima T, Kawazu C, Kitazume Y, Kojima M, Miura H, Murakami K, Murata M, Ninomiya N, Nishiyori H, Noma S, Ogawa C, Sano T, Simon C, Tagami M, Takahashi Y, Kawai J, Hayashizaki Y. The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. *Nature genetics*. 2009;41(5):553-62. Epub 2009/04/21. doi: 10.1038/ng.375. PubMed PMID: 19377474.
 156. Culhane AC, **Quackenbush J**. Confounding effects in "A six-gene signature predicting breast cancer lung metastasis". *Cancer research*. 2009;69(18):7480-5. Epub 2009/09/03. doi: 10.1158/0008-5472.CAN-08-3350. PubMed PMID: 19723662; PubMed Central PMCID: PMC3128918.
 157. Dutta B, Kanani H, **Quackenbush J**, Klapa MI. Time-series integrated "omic" analyses to elucidate short-term stress-induced responses in plant liquid cultures. *Biotechnology and bioengineering*. 2009;102(1):264-79. Epub 2008/10/30. doi: 10.1002/bit.22036. PubMed PMID: 18958862; PubMed Central PMCID: PMC4134133.
 158. Fedorov VB, Goropashnaya AV, Toien O, Stewart NC, Gracey AY, Chang C, Qin S, Perteu G, **Quackenbush J**, Showe LC, Showe MK, Boyer BB, Barnes BM. Elevated expression of protein biosynthesis genes in liver and muscle of hibernating black bears (*Ursus americanus*). *Physiological genomics*. 2009;37(2):108-18. Epub 2009/02/26. doi: 10.1152/physiolgenomics.90398.2008. PubMed PMID: 19240299; PubMed Central PMCID: PMC3774579.
 159. Gorgun G, Ramsay AG, Holderried TA, Zahrieh D, Le Dieu R, Liu F, **Quackenbush J**, Croce CM, Gribben JG. E(mu)-TCL1 mice represent a model for immunotherapeutic reversal of chronic lymphocytic leukemia-induced T-cell

- dysfunction. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(15):6250-5. Epub 2009/04/01. doi: 10.1073/pnas.0901166106. PubMed PMID: 19332800; PubMed Central PMCID: PMC2669383.
160. Hellner K, Mar J, Fang F, **Quackenbush J**, Munger K. HPV16 E7 oncogene expression in normal human epithelial cells causes molecular changes indicative of an epithelial to mesenchymal transition. Virology. 2009;391(1):57-63. Epub 2009/06/26. doi: 10.1016/j.virol.2009.05.036. PubMed PMID: 19552933; PubMed Central PMCID: PMC2728214.
 161. Mar JC, Kimura Y, Schroder K, Irvine KM, Hayashizaki Y, Suzuki H, Hume D, **Quackenbush J**. Data-driven normalization strategies for high-throughput quantitative RT-PCR. BMC bioinformatics. 2009;10:110. Epub 2009/04/21. doi: 10.1186/1471-2105-10-110. PubMed PMID: 19374774; PubMed Central PMCID: PMC2680405.
 162. Mar JC, **Quackenbush J**. Decomposition of gene expression state space trajectories. PLoS computational biology. 2009;5(12):e1000626. Epub 2009/12/31. doi: 10.1371/journal.pcbi.1000626. PubMed PMID: 20041215; PubMed Central PMCID: PMC2791157.
 163. **Quackenbush J**. Data reporting standards: making the things we use better. Genome medicine. 2009;1(11):111. Epub 2009/11/27. doi: 10.1186/gm111. PubMed PMID: 19939294; PubMed Central PMCID: PMC2808746.
 164. Santagata S, Maire CL, Idbaih A, Geffers L, Correll M, Holton K, **Quackenbush J**, Ligon KL. CRX is a diagnostic marker of retinal and pineal lineage tumors. PloS one. 2009;4(11):e7932. Epub 2009/11/26. doi: 10.1371/journal.pone.0007932. PubMed PMID: 19936203; PubMed Central PMCID: PMC2775954.
 165. Speers C, Tsimelzon A, Sexton K, Herrick AM, Gutierrez C, Culhane A, **Quackenbush J**, Hilsenbeck S, Chang J, Brown P. Identification of novel kinase targets for the treatment of estrogen receptor-negative breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(20):6327-40. Epub 2009/10/08. doi: 10.1158/1078-0432.CCR-09-1107. PubMed PMID: 19808870; PubMed Central PMCID: PMC2763053.
 166. Antonescu C, Antonescu V, Sultana R, **Quackenbush J**. Using the DFCI gene index databases for biological discovery. Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]. 2010;Chapter 1:Unit1 6 1-36. Epub 2010/03/06. doi: 10.1002/0471250953.bi0106s29. PubMed PMID: 20205187; PubMed Central PMCID: PMC4205239.
 167. Chen DT, Nasir A, Culhane A, Venkataramu C, Fulp W, Rubio R, Wang T, Agrawal D, McCarthy SM, Gruidl M, Bloom G, Anderson T, White J, **Quackenbush J**, Yeatman T. Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. Breast cancer research and treatment. 2010;119(2):335-46. Epub 2009/03/07. doi: 10.1007/s10549-009-0344-y. PubMed PMID: 19266279; PubMed Central PMCID: PMC2796276.
 168. Chittenden TW, Pak J, Rubio R, Cheng H, Holton K, Prendergast N, Glinskii V, Cai Y, Culhane A, Bentink S, Schwede M, Mar JC, Howe EA, Aryee M, Sultana R, Lanahan AA, Taylor JM, Holmes C, Hahn WC, Zhao JJ, Iglehart JD, **Quackenbush J**. Therapeutic implications of GIPC1 silencing in cancer. PloS one. 2010;5(12):e15581. Epub 2011/01/07. doi: 10.1371/journal.pone.0015581. PubMed PMID: 21209904; PubMed Central PMCID: PMC3012716.

169. Colak D, Chishti MA, Al-Bakheet AB, Al-Qahtani A, Shoukri MM, Goyns MH, Ozand PT, **Quackenbush J**, Park BH, Kaya N. Integrative and comparative genomics analysis of early hepatocellular carcinoma differentiated from liver regeneration in young and old. *Molecular cancer*. 2010;9:146. Epub 2010/06/15. doi: 10.1186/1476-4598-9-146. PubMed PMID: 20540791; PubMed Central PMCID: PMC2898705.
170. Culhane AC, Schwarzl T, Sultana R, Picard KC, Picard SC, Lu TH, Franklin KR, French SJ, Papenhausen G, Correll M, **Quackenbush J**. GeneSigDB--a curated database of gene expression signatures. *Nucleic acids research*. 2010;38(Database issue):D716-25. Epub 2009/11/26. doi: 10.1093/nar/gkp1015. PubMed PMID: 19934259; PubMed Central PMCID: PMC2808880.
171. Tanaka N, Huttenhower C, Nosh K, Baba Y, Shima K, **Quackenbush J**, Haigis KM, Giovannucci E, Fuchs CS, Ogino S. Novel application of structural equation modeling to correlation structure analysis of CpG island methylation in colorectal cancer. *The American journal of pathology*. 2010;177(6):2731-40. Epub 2010/11/03. doi: 10.2353/ajpath.2010.100361. PubMed PMID: 21037082; PubMed Central PMCID: PMC2993288.
172. Tseveleki V, Rubio R, Vamvakas SS, White J, Taoufik E, Petit E, **Quackenbush J**, Probert L. Comparative gene expression analysis in mouse models for multiple sclerosis, Alzheimer's disease and stroke for identifying commonly regulated and disease-specific gene changes. *Genomics*. 2010;96(2):82-91. Epub 2010/05/04. doi: 10.1016/j.ygeno.2010.04.004. PubMed PMID: 20435134; PubMed Central PMCID: PMC4205236.
173. Bontempi G, Haibe-Kains B, Desmedt C, Sotiriou C, **Quackenbush J**. Multiple-input multiple-output causal strategies for gene selection. *BMC bioinformatics*. 2011;12:458. Epub 2011/11/29. doi: 10.1186/1471-2105-12-458. PubMed PMID: 22118187; PubMed Central PMCID: PMC3323860.
174. Cannon EK, Birkett SM, Braun BL, Kodavali S, Jennewein DM, Yilmaz A, Antonescu V, Antonescu C, Harper LC, Gardiner JM, Schaeffer ML, Campbell DA, Andorf CM, Andorf D, Lisch D, Koch KE, McCarty DR, **Quackenbush J**, Grotewold E, Lushbough CM, Sen TZ, Lawrence CJ. POPcorn: An Online Resource Providing Access to Distributed and Diverse Maize Project Data. *International journal of plant genomics*. 2011;2011:923035. Epub 2012/01/19. doi: 10.1155/2011/923035. PubMed PMID: 22253616; PubMed Central PMCID: PMC3255282.
175. Chervitz SA, Deutsch EW, Field D, Parkinson H, **Quackenbush J**, Rocca-Serra P, Sansone SA, Stoeckert CJ, Jr., Taylor CF, Taylor R, Ball CA. Data standards for Omics data: the basis of data sharing and reuse. *Methods Mol Biol*. 2011;719:31-69. Epub 2011/03/04. doi: 10.1007/978-1-61779-027-0_2. PubMed PMID: 21370078; PubMed Central PMCID: PMC4152841.
176. Gewurz BE, Mar JC, Padi M, Zhao B, Shinnars NP, Takasaki K, Bedoya E, Zou JY, Cahir-McFarland E, **Quackenbush J**, Kieff E. Canonical NF-kappaB activation is essential for Epstein-Barr virus latent membrane protein 1 TES2/CTAR2 gene regulation. *Journal of virology*. 2011;85(13):6764-73. Epub 2011/05/06. doi: 10.1128/JVI.00422-11. PubMed PMID: 21543491; PubMed Central PMCID: PMC3126495.
177. Howe EA, Sinha R, Schlauch D, **Quackenbush J**. RNA-Seq analysis in MeV. *Bioinformatics*. 2011;27(22):3209-10. Epub 2011/10/07. doi:

- 10.1093/bioinformatics/btr490. PubMed PMID: 21976420; PubMed Central PMCID: PMC3208390.
178. Jirawatnotai S, Hu Y, Michowski W, Elias JE, Becks L, Bienvenu F, Zagazdzon A, Goswami T, Wang YE, Clark AB, Kunkel TA, van Harn T, Xia B, Correll M, **Quackenbush J**, Livingston DM, Gygi SP, Sicinski P. A function for cyclin D1 in DNA repair uncovered by protein interactome analyses in human cancers. *Nature*. 2011;474(7350):230-4. Epub 2011/06/10. doi: 10.1038/nature10155. PubMed PMID: 21654808; PubMed Central PMCID: PMC3134411.
 179. Konstantinopoulos PA, Cannistra SA, Fountzilas H, Culhane A, Pillay K, Rueda B, Cramer D, Seiden M, Birrer M, Coukos G, Zhang L, **Quackenbush J**, Spentzos D. Integrated analysis of multiple microarray datasets identifies a reproducible survival predictor in ovarian cancer. *PloS one*. 2011;6(3):e18202. Epub 2011/04/12. doi: 10.1371/journal.pone.0018202. PubMed PMID: 21479231; PubMed Central PMCID: PMC3066217.
 180. Liu F, White JA, Antonescu C, Gusenleitner D, **Quackenbush J**. GCOD - GeneChip Oncology Database. *BMC bioinformatics*. 2011;12:46. Epub 2011/02/05. doi: 10.1186/1471-2105-12-46. PubMed PMID: 21291543; PubMed Central PMCID: PMC3045303.
 181. Mar JC, Matigian NA, Mackay-Sim A, Mellick GD, Sue CM, Silburn PA, McGrath JJ, **Quackenbush J**, Wells CA. Variance of gene expression identifies altered network constraints in neurological disease. *PLoS genetics*. 2011;7(8):e1002207. Epub 2011/08/20. doi: 10.1371/journal.pgen.1002207. PubMed PMID: 21852951; PubMed Central PMCID: PMC3154954.
 182. Mar JC, Matigian NA, **Quackenbush J**, Wells CA. attract: A method for identifying core pathways that define cellular phenotypes. *PloS one*. 2011;6(10):e25445. Epub 2011/10/25. doi: 10.1371/journal.pone.0025445. PubMed PMID: 22022396; PubMed Central PMCID: PMC3194807.
 183. Mar JC, Wells CA, **Quackenbush J**. Defining an informativeness metric for clustering gene expression data. *Bioinformatics*. 2011;27(8):1094-100. Epub 2011/02/19. doi: 10.1093/bioinformatics/btr074. PubMed PMID: 21330289; PubMed Central PMCID: PMC3072547.
 184. Marko NF, **Quackenbush J**, Weil RJ. Why is there a lack of consensus on molecular subgroups of glioblastoma? Understanding the nature of biological and statistical variability in glioblastoma expression data. *PloS one*. 2011;6(7):e20826. Epub 2011/08/11. doi: 10.1371/journal.pone.0020826. PubMed PMID: 21829433; PubMed Central PMCID: PMC3145641.
 185. **Quackenbush J**. The Human Genome: Book of Essential Knowledge (Curiosity Guides): Imagine! Publishing, An Imprint of Charlesbridge; 2011. 176 p.
 186. Sathirapongsasuti JF, Lee H, Horst BA, Brunner G, Cochran AJ, Binder S, **Quackenbush J**, Nelson SF. Exome sequencing-based copy-number variation and loss of heterozygosity detection: ExomeCNV. *Bioinformatics*. 2011;27(19):2648-54. Epub 2011/08/11. doi: 10.1093/bioinformatics/btr462. PubMed PMID: 21828086; PubMed Central PMCID: PMC3179661.
 187. Schroder MS, Culhane AC, **Quackenbush J**, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics*. 2011;27(22):3206-8. Epub 2011/09/10. doi: 10.1093/bioinformatics/btr511. PubMed PMID: 21903630; PubMed Central PMCID: PMC3208391.

188. Zhao B, Mar JC, Maruo S, Lee S, Gewurz BE, Johannsen E, Holton K, Rubio R, Takada K, **Quackenbush J**, Kieff E. Epstein-Barr virus nuclear antigen 3C regulated genes in lymphoblastoid cell lines. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(1):337-42. Epub 2010/12/22. doi: 10.1073/pnas.1017419108. PubMed PMID: 21173222; PubMed Central PMCID: PMC3017193.
189. Zhao B, Zou J, Wang H, Johannsen E, Peng CW, **Quackenbush J**, Mar JC, Morton CC, Freedman ML, Blacklow SC, Aster JC, Bernstein BE, Kieff E. Epstein-Barr virus exploits intrinsic B-lymphocyte transcription programs to achieve immortal cell growth. *Proceedings of the National Academy of Sciences of the United States of America*. 2011;108(36):14902-7. Epub 2011/07/13. doi: 10.1073/pnas.1108892108. PubMed PMID: 21746931; PubMed Central PMCID: PMC3169132.
190. Bentink S, Haibe-Kains B, Risch T, Fan JB, Hirsch MS, Holton K, Rubio R, April C, Chen J, Wickham-Garcia E, Liu J, Culhane A, Drapkin R, **Quackenbush J**, Matulonis UA. Angiogenic mRNA and microRNA gene expression signature predicts a novel subtype of serous ovarian cancer. *PloS one*. 2012;7(2):e30269. Epub 2012/02/22. doi: 10.1371/journal.pone.0030269. PubMed PMID: 22348002; PubMed Central PMCID: PMC3278409.
191. Chittenden TW, Howe EA, Taylor JM, Mar JC, Aryee MJ, Gomez H, Sultana R, Braisted J, Nair SJ, **Quackenbush J**, Holmes C. nEASE: a method for gene ontology subclassification of high-throughput gene expression data. *Bioinformatics*. 2012;28(5):726-8. Epub 2012/01/17. doi: 10.1093/bioinformatics/bts011. PubMed PMID: 22247278.
192. Culhane AC, Schroder MS, Sultana R, Picard SC, Martinelli EN, Kelly C, Haibe-Kains B, Kapushesky M, St Pierre AA, Flahive W, Picard KC, Gusenleitner D, Papenhausen G, O'Connor N, Correll M, **Quackenbush J**. GeneSigDB: a manually curated database and resource for analysis of gene expression signatures. *Nucleic acids research*. 2012;40(Database issue):D1060-6. Epub 2011/11/24. doi: 10.1093/nar/gkr901. PubMed PMID: 22110038; PubMed Central PMCID: PMC3245038.
193. Desmedt C, Majjaj S, Kheddoumi N, Singhal SK, Haibe-Kains B, El Ouriaghli F, Chaboteaux C, Michiels S, Lallemand F, Journe F, Duvillier H, Loi S, **Quackenbush J**, Dekoninck S, Blanpain C, Lagneaux L, Houhou N, Delorenzi M, Larsimont D, Piccart M, Sotiriou C. Characterization and clinical evaluation of CD10+ stroma cells in the breast cancer microenvironment. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012;18(4):1004-14. Epub 2012/01/12. doi: 10.1158/1078-0432.CCR-11-0383. PubMed PMID: 22235100.
194. Fountzilias E, Kelly AD, Perez-Atayde AR, Goldsmith J, Konstantinopoulos PA, Francoeur N, Correll M, Rubio R, Hu L, Gebhardt MC, **Quackenbush J**, Spentzos D. A microRNA activity map of human mesenchymal tumors: connections to oncogenic pathways; an integrative transcriptomic study. *BMC genomics*. 2012;13:332. Epub 2012/07/25. doi: 10.1186/1471-2164-13-332. PubMed PMID: 22823907; PubMed Central PMCID: PMC3443663.
195. Gewurz BE, Towfic F, Mar JC, Shinnars NP, Takasaki K, Zhao B, Cahir-McFarland ED, **Quackenbush J**, Xavier RJ, Kieff E. Genome-wide siRNA screen for mediators of NF-kappaB activation. *Proceedings of the National Academy of*

- Sciences of the United States of America. 2012;109(7):2467-72. Epub 2012/02/07. doi: 10.1073/pnas.1120542109. PubMed PMID: 22308454; PubMed Central PMCID: PMC3289371.
196. Gulbahce N, Yan H, Dricot A, Padi M, Byrdsong D, Franchi R, Lee DS, Rozenblatt-Rosen O, Mar JC, Calderwood MA, Baldwin A, Zhao B, Santhanam B, Braun P, Simonis N, Huh KW, Hellner K, Grace M, Chen A, Rubio R, Marto JA, Christakis NA, Kieff E, Roth FP, Roeklein-Canfield J, Decaprio JA, Cusick ME, **Quackenbush J**, Hill DE, Munger K, Vidal M, Barabasi AL. Viral perturbations of host networks reflect disease etiology. PLoS computational biology. 2012;8(6):e1002531. Epub 2012/07/05. doi: 10.1371/journal.pcbi.1002531. PubMed PMID: 22761553; PubMed Central PMCID: PMC3386155.
 197. Gusenleitner D, Howe EA, Bentink S, **Quackenbush J**, Culhane AC. iBBiG: iterative binary bi-clustering of gene sets. Bioinformatics. 2012;28(19):2484-92. Epub 2012/07/14. doi: 10.1093/bioinformatics/bts438. PubMed PMID: 22789589; PubMed Central PMCID: PMC3463116.
 198. Haibe-Kains B, Desmedt C, Loi S, Culhane AC, Bontempi G, **Quackenbush J**, Sotiriou C. A three-gene model to robustly identify breast cancer molecular subtypes. Journal of the National Cancer Institute. 2012;104(4):311-25. Epub 2012/01/21. doi: 10.1093/jnci/djr545. PubMed PMID: 22262870; PubMed Central PMCID: PMC3283537.
 199. Haibe-Kains B, Olsen C, Djebbari A, Bontempi G, Correll M, Bouton C, **Quackenbush J**. Predictive networks: a flexible, open source, web application for integration and analysis of human gene networks. Nucleic acids research. 2012;40(Database issue):D866-75. Epub 2011/11/19. doi: 10.1093/nar/gkr1050. PubMed PMID: 22096235; PubMed Central PMCID: PMC3245161.
 200. Milbury CA, Correll M, **Quackenbush J**, Rubio R, Makrigiorgos GM. COLD-PCR enrichment of rare cancer mutations prior to targeted amplicon resequencing. Clinical chemistry. 2012;58(3):580-9. Epub 2011/12/24. doi: 10.1373/clinchem.2011.176198. PubMed PMID: 22194627; PubMed Central PMCID: PMC3418918.
 201. Rozenblatt-Rosen O, Deo RC, Padi M, Adelmant G, Calderwood MA, Rolland T, Grace M, Dricot A, Askenazi M, Tavares M, Pevzner SJ, Abderazzaq F, Byrdsong D, Carvunis AR, Chen AA, Cheng J, Correll M, Duarte M, Fan C, Feltkamp MC, Ficarro SB, Franchi R, Garg BK, Gulbahce N, Hao T, Holthaus AM, James R, Korkhin A, Litovchick L, Mar JC, Pak TR, Rabello S, Rubio R, Shen Y, Singh S, Spangle JM, Tasan M, Wanamaker S, Webber JT, Roeklein-Canfield J, Johannsen E, Barabasi AL, Beroukhim R, Kieff E, Cusick ME, Hill DE, Munger K, Marto JA, **Quackenbush J**, Roth FP, DeCaprio JA, Vidal M. Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. Nature. 2012;487(7408):491-5. Epub 2012/07/20. doi: 10.1038/nature11288. PubMed PMID: 22810586; PubMed Central PMCID: PMC3408847.
 202. Skiadas CC, Duan S, Correll M, Rubio R, Karaca N, Ginsburg ES, **Quackenbush J**, Racowsky C. Ovarian reserve status in young women is associated with altered gene expression in membrana granulosa cells. Molecular human reproduction. 2012;18(7):362-71. Epub 2012/02/23. doi: 10.1093/molehr/gas008. PubMed PMID: 22355044; PubMed Central PMCID: PMC3378309.
 203. Vidal M, Chan DW, Gerstein M, Mann M, Omenn GS, Tagle D, Sechi S, Apweiler R, Bader J, Barnes S, Bettauer R, Chan MM, Coon J, Dolinski K, Dowell BL,

- Edmonds CG, Gaudet S, Gerszten R, Kagan J, Mazurchuk R, McShane L, Moritz RL, Palmer-Toy D, Pandey A, Polanski M, **Quackenbush J**, Sheeley D, Skates S, Srivastava S, Wang W, Weng Z, White F. The human proteome - a scientific opportunity for transforming diagnostics, therapeutics, and healthcare. *Clinical proteomics*. 2012;9(1):6. Epub 2012/05/16. doi: 10.1186/1559-0275-9-6. PubMed PMID: 22583803; PubMed Central PMCID: PMC3388576.
204. Waldron L, Ogino S, Hoshida Y, Shima K, McCart Reed AE, Simpson PT, Baba Y, Nosh K, Segata N, Vargas AC, Cummings MC, Lakhani SR, Kirkner GJ, Giovannucci E, **Quackenbush J**, Golub TR, Fuchs CS, Parmigiani G, Huttenhower C. Expression profiling of archival tumors for long-term health studies. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012;18(22):6136-46. Epub 2012/11/09. doi: 10.1158/1078-0432.CCR-12-1915. PubMed PMID: 23136189; PubMed Central PMCID: PMC3500412.
205. Wang ZC, Birkbak NJ, Culhane AC, Drapkin R, Fatima A, Tian R, Schwede M, Alsop K, Daniels KE, Piao H, Liu J, Etemadmoghadam D, Miron A, Salvesen HB, Mitchell G, DeFazio A, **Quackenbush J**, Berkowitz RS, Iglehart JD, Bowtell DD, Australian Ovarian Cancer Study G, Matulonis UA. Profiles of genomic instability in high-grade serous ovarian cancer predict treatment outcome. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2012;18(20):5806-15. Epub 2012/08/23. doi: 10.1158/1078-0432.CCR-12-0857. PubMed PMID: 22912389; PubMed Central PMCID: PMC4205235.
206. Beck AH, Knoblauch NW, Hefti MM, Kaplan J, Schnitt SJ, Culhane AC, Schroeder MS, Risch T, **Quackenbush J**, Haibe-Kains B. Significance analysis of prognostic signatures. *PLoS computational biology*. 2013;9(1):e1002875. Epub 2013/02/01. doi: 10.1371/journal.pcbi.1002875. PubMed PMID: 23365551; PubMed Central PMCID: PMC3554539.
207. Correll M, Johnson CK, Ferrari G, Brizzio M, Mak AW, **Quackenbush J**, Shaw RE, Zapolanski A, Grau JB. Mutational analysis clopidogrel resistance and platelet function in patients scheduled for coronary artery bypass grafting. *Genomics*. 2013;101(6):313-7. Epub 2013/03/07. doi: 10.1016/j.ygeno.2013.01.005. PubMed PMID: 23462555; PubMed Central PMCID: PMC4149181.
208. De Rienzo A, Richards WG, Yeap BY, Coleman MH, Sugarbaker PE, Chirieac LR, Wang YE, **Quackenbush J**, Jensen RV, Bueno R. Sequential binary gene ratio tests define a novel molecular diagnostic strategy for malignant pleural mesothelioma. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2013;19(9):2493-502. Epub 2013/03/16. doi: 10.1158/1078-0432.CCR-12-2117. PubMed PMID: 23493352; PubMed Central PMCID: PMC3644001.
209. Glass K, Huttenhower C, **Quackenbush J**, Yuan GC. Passing messages between biological networks to refine predicted interactions. *PloS one*. 2013;8(5):e64832. Epub 2013/06/07. doi: 10.1371/journal.pone.0064832. PubMed PMID: 23741402; PubMed Central PMCID: PMC3669401.
210. Haibe-Kains B, El-Hachem N, Birkbak NJ, Jin AC, Beck AH, Aerts HJ, **Quackenbush J**. Inconsistency in large pharmacogenomic studies. *Nature*. 2013;504(7480):389-93. Epub 2013/11/29. doi: 10.1038/nature12831. PubMed PMID: 24284626; PubMed Central PMCID: PMC4237165.
211. Kelly AD, Haibe-Kains B, Janeway KA, Hill KE, Howe E, Goldsmith J, Kurek K, Perez-Atayde AR, Francoeur N, Fan JB, April C, Schneider H, Gebhardt MC,

- Culhane A, **Quackenbush J**, Spentzos D. MicroRNA paraffin-based studies in osteosarcoma reveal reproducible independent prognostic profiles at 14q32. *Genome medicine*. 2013;5(1):2. Epub 2013/01/24. doi: 10.1186/gm406. PubMed PMID: 23339462; PubMed Central PMCID: PMC3706900.
212. Kelly AD, Hill KE, Correll M, Hu L, Wang YE, Rubio R, Duan S, **Quackenbush J**, Spentzos D. Next-generation sequencing and microarray-based interrogation of microRNAs from formalin-fixed, paraffin-embedded tissue: preliminary assessment of cross-platform concordance. *Genomics*. 2013;102(1):8-14. Epub 2013/04/09. doi: 10.1016/j.ygeno.2013.03.008. PubMed PMID: 23562991; PubMed Central PMCID: PMC4116671.
213. Lam HC, Cloonan SM, Bhashyam AR, Haspel JA, Singh A, Sathirapongsasuti JF, Cervo M, Yao H, Chung AL, Mizumura K, An CH, Shan B, Franks JM, Haley KJ, Owen CA, Tesfaigzi Y, Washko GR, **Quackenbush J**, Silverman EK, Rahman I, Kim HP, Mahmood A, Biswal SS, Ryter SW, Choi AM. Histone deacetylase 6-mediated selective autophagy regulates COPD-associated cilia dysfunction. *The Journal of clinical investigation*. 2013;123(12):5212-30. Epub 2013/11/10. doi: 10.1172/JCI69636. PubMed PMID: 24200693; PubMed Central PMCID: PMC3859407.
214. Larson JL, Huttenhower C, **Quackenbush J**, Yuan GC. A tiered hidden Markov model characterizes multi-scale chromatin states. *Genomics*. 2013;102(1):1-7. Epub 2013/04/11. doi: 10.1016/j.ygeno.2013.03.009. PubMed PMID: 23570996; PubMed Central PMCID: PMC3676702.
215. Merritt MA, Bentink S, Schwede M, Iwanicki MP, **Quackenbush J**, Woo T, Agoston ES, Reinhardt F, Crum CP, Berkowitz RS, Mok SC, Witt AE, Jones MA, Wang B, Ince TA. Gene expression signature of normal cell-of-origin predicts ovarian tumor outcomes. *PloS one*. 2013;8(11):e80314. Epub 2013/12/05. doi: 10.1371/journal.pone.0080314. PubMed PMID: 24303006; PubMed Central PMCID: PMC3841174.
216. Papillon-Cavanagh S, De Jay N, Hachem N, Olsen C, Bontempi G, Aerts HJ, **Quackenbush J**, Haibe-Kains B. Comparison and validation of genomic predictors for anticancer drug sensitivity. *Journal of the American Medical Informatics Association : JAMIA*. 2013;20(4):597-602. Epub 2013/01/29. doi: 10.1136/amiajnl-2012-001442. PubMed PMID: 23355484; PubMed Central PMCID: PMC3721163.
217. Portal D, Zhou H, Zhao B, Kharchenko PV, Lowry E, Wong L, **Quackenbush J**, Holloway D, Jiang S, Lu Y, Kieff E. Epstein-Barr virus nuclear antigen leader protein localizes to promoters and enhancers with cell transcription factors and EBNA2. *Proceedings of the National Academy of Sciences of the United States of America*. 2013;110(46):18537-42. Epub 2013/10/30. doi: 10.1073/pnas.1317608110. PubMed PMID: 24167291; PubMed Central PMCID: PMC3832032.
218. Schroder MS, Gusenleitner D, **Quackenbush J**, Culhane AC, Haibe-Kains B. RamiGO: an R/Bioconductor package providing an AmiGO visualize interface. *Bioinformatics*. 2013;29(5):666-8. Epub 2013/01/09. doi: 10.1093/bioinformatics/bts708. PubMed PMID: 23297033; PubMed Central PMCID: PMC3582261.
219. Schwede M, Spentzos D, Bentink S, Hofmann O, Haibe-Kains B, Harrington D, **Quackenbush J**, Culhane AC. Stem cell-like gene expression in ovarian cancer predicts type II subtype and prognosis. *PloS one*. 2013;8(3):e57799. Epub

- 2013/03/29. doi: 10.1371/journal.pone.0057799. PubMed PMID: 23536770; PubMed Central PMCID: PMC3594231.
220. Zhou X, Qiu W, Sathirapongsasuti JF, Cho MH, Mancini JD, Lao T, Thibault DM, Litonjua AA, Bakke PS, Gulsvik A, Lomas DA, Beaty TH, Hersh CP, Anderson C, Geigenmuller U, Raby BA, Rennard SI, Perrella MA, Choi AM, **Quackenbush J**, Silverman EK. Gene expression analysis uncovers novel hedgehog interacting protein (HHIP) effects in human bronchial epithelial cells. *Genomics*. 2013;101(5):263-72. Epub 2013/03/06. doi: 10.1016/j.ygeno.2013.02.010. PubMed PMID: 23459001; PubMed Central PMCID: PMC3659826.
 221. Adamia S, Bar-Natan M, Haibe-Kains B, Pilarski PM, Bach C, Pevzner S, Calimeri T, Avet-Loiseau H, Lode L, Verselis S, Fox EA, Galinsky I, Mathews S, Dagogo-Jack I, Wadleigh M, Steensma DP, Motyckova G, Deangelo DJ, **Quackenbush J**, Tenen DG, Stone RM, Griffin JD. NOTCH2 and FLT3 gene mis-splicings are common events in patients with acute myeloid leukemia (AML): new potential targets in AML. *Blood*. 2014;123(18):2816-25. Epub 2014/02/28. doi: 10.1182/blood-2013-02-481507. PubMed PMID: 24574459; PubMed Central PMCID: PMC4007608.
 222. Adamia S, Haibe-Kains B, Pilarski PM, Bar-Natan M, Pevzner S, Avet-Loiseau H, Lode L, Verselis S, Fox EA, Burke J, Galinsky I, Dagogo-Jack I, Wadleigh M, Steensma DP, Motyckova G, Deangelo DJ, **Quackenbush J**, Stone R, Griffin JD. A genome-wide aberrant RNA splicing in patients with acute myeloid leukemia identifies novel potential disease markers and therapeutic targets. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2014;20(5):1135-45. Epub 2013/11/29. doi: 10.1158/1078-0432.CCR-13-0956. PubMed PMID: 24284058.
 223. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Cavalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebbers F, Rietbergen MM, Leemans CR, Dekker A, **Quackenbush J**, Gillies RJ, Lambin P. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications*. 2014;5:4006. Epub 2014/06/04. doi: 10.1038/ncomms5006. PubMed PMID: 24892406; PubMed Central PMCID: PMC4059926.
 224. Chu JH, Hersh CP, Castaldi PJ, Cho MH, Raby BA, Laird N, Bowler R, Rennard S, Loscalzo J, **Quackenbush J**, Silverman EK. Analyzing networks of phenotypes in complex diseases: methodology and applications in COPD. *BMC systems biology*. 2014;8:78. Epub 2014/06/27. doi: 10.1186/1752-0509-8-78. PubMed PMID: 24964944; PubMed Central PMCID: PMC4105829.
 225. Ferrari G, **Quackenbush J**, Strobeck J, Hu L, Johnson CK, Mak A, Shaw RE, Sayles K, Brizzio ME, Zapolanski A, Grau JB. Comparative genome-wide transcriptional analysis of human left and right internal mammary arteries. *Genomics*. 2014;104(1):36-44. Epub 2014/05/27. doi: 10.1016/j.ygeno.2014.04.005. PubMed PMID: 24858532; PubMed Central PMCID: PMC4152843.
 226. Glass K, **Quackenbush J**, Silverman EK, Celli B, Rennard SI, Yuan GC, DeMeo DL. Sexually-dimorphic targeting of functionally-related genes in COPD. *BMC systems biology*. 2014;8:118. Epub 2014/11/29. doi: 10.1186/s12918-014-0118-y. PubMed PMID: 25431000; PubMed Central PMCID: PMC4269917.
 227. Hatzis C, Bedard PL, Birkbak NJ, Beck AH, Aerts HJ, Stem DF, Shi L, Clarke R, **Quackenbush J**, Haibe-Kains B. Enhancing reproducibility in cancer drug

- screening: how do we move forward? *Cancer research*. 2014;74(15):4016-23. Epub 2014/07/13. doi: 10.1158/0008-5472.CAN-14-0725. PubMed PMID: 25015668; PubMed Central PMCID: PMC4119520.
228. Jirawatnotai S, Sharma S, Michowski W, Suktitipat B, Geng Y, **Quackenbush J**, Elias JE, Gygi SP, Wang YE, Sicinski P. The cyclin D1-CDK4 oncogenic interactome enables identification of potential novel oncogenes and clinical prognosis. *Cell Cycle*. 2014;13(18):2889-900. Epub 2014/12/09. doi: 10.4161/15384101.2014.946850. PubMed PMID: 25486477.
 229. Lee JW, Norden AD, Ligon KL, Golby AJ, Beroukhir R, **Quackenbush J**, Wells W, Oelschlager K, Maetzold D, Wen PY. Tumor associated seizures in glioblastomas are influenced by survival gene expression in a region-specific manner: a gene expression imaging study. *Epilepsy research*. 2014;108(5):843-52. Epub 2014/04/03. doi: 10.1016/j.eplepsyres.2014.02.021. PubMed PMID: 24690158; PubMed Central PMCID: PMC4076930.
 230. Li WQ, Han J, Widlund HR, Correll M, Wang YE, **Quackenbush J**, Mihm MC, Canales AL, Wu S, Golub T, Hoshida Y, Hunter DJ, Murphy G, Kupper TS, Qureshi AA. CXCR4 pathway associated with family history of melanoma. *Cancer causes & control : CCC*. 2014;25(1):125-32. Epub 2013/10/26. doi: 10.1007/s10552-013-0315-9. PubMed PMID: 24158781; PubMed Central PMCID: PMC4100594.
 231. Mason EA, Mar JC, Laslett AL, Pera MF, **Quackenbush J**, Wolvetang E, Wells CA. Gene expression variability as a unifying element of the pluripotency network. *Stem cell reports*. 2014;3(2):365-77. Epub 2014/09/26. doi: 10.1016/j.stemcr.2014.06.008. PubMed PMID: 25254348; PubMed Central PMCID: PMC4175554.
 232. Olsen C, Bontempi G, Emmert-Streib F, **Quackenbush J**, Haibe-Kains B. Relevance of different prior knowledge sources for inferring gene interaction networks. *Frontiers in genetics*. 2014;5:177. Epub 2014/07/11. doi: 10.3389/fgene.2014.00177. PubMed PMID: 25009552; PubMed Central PMCID: PMC4067568.
 233. Olsen C, Fleming K, Prendergast N, Rubio R, Emmert-Streib F, Bontempi G, Haibe-Kains B, **Quackenbush J**. Inference and validation of predictive gene networks from biomedical literature and gene expression data. *Genomics*. 2014;103(5-6):329-36. Epub 2014/04/03. doi: 10.1016/j.ygeno.2014.03.004. PubMed PMID: 24691108; PubMed Central PMCID: PMC4119824.
 234. **Quackenbush J**. Learning to share. *Scientific American*. 2014;311(1):S22. Epub 2014/07/01. PubMed PMID: 24974706.
 235. **Quackenbush J**. Perspective: Learning to share. *Nature*. 2014;509(7502):S68. Epub 2014/05/30. doi: 10.1038/509S68a. PubMed PMID: 24870827; PubMed Central PMCID: PMC4205237.
 236. Sharron Lin X, Hu L, Sandy K, Correll M, **Quackenbush J**, Wu CL, Scott McDougal W. Differentiating progressive from nonprogressive T1 bladder cancer by gene expression profiling: applying RNA-sequencing analysis on archived specimens. *Urologic oncology*. 2014;32(3):327-36. Epub 2013/09/24. doi: 10.1016/j.urolonc.2013.06.014. PubMed PMID: 24055427.
 237. Yang IV, Pedersen BS, Rabinovich E, Hennessy CE, Davidson EJ, Murphy E, Guardela BJ, Tedrow JR, Zhang Y, Singh MK, Correll M, Schwarz MI, Geraci M, Sciurba FC, **Quackenbush J**, Spira A, Kaminski N, Schwartz DA. Relationship of

- DNA methylation and gene expression in idiopathic pulmonary fibrosis. *American journal of respiratory and critical care medicine*. 2014;190(11):1263-72. Epub 2014/10/22. doi: 10.1164/rccm.201408-1452OC. PubMed PMID: 25333685; PubMed Central PMCID: PMC4315819.
238. Campbell JD, Liu G, Luo L, Xiao J, Gerrein J, Juan-Guardela B, Tedrow J, Alekseyev YO, Yang IV, Correll M, Geraci M, **Quackenbush J**, Sciruba F, Schwartz DA, Kaminski N, Johnson WE, Monti S, Spira A, Beane J, Lenburg ME. Assessment of microRNA differential expression and detection in multiplexed small RNA sequencing data. *RNA*. 2015;21(2):164-71. Epub 2014/12/19. doi: 10.1261/rna.046060.114. PubMed PMID: 25519487; PubMed Central PMCID: PMC4338344.
 239. Castaldi PJ, Cho MH, Zhou X, Qiu W, McGeachie M, Celli B, Bakke P, Gulsvik A, Lomas DA, Crapo JD, Beaty TH, Rennard S, Harshfield B, Lange C, Singh D, Tal-Singer R, Riley JH, **Quackenbush J**, Raby BA, Carey VJ, Silverman EK, Hersh CP. Genetic control of gene expression at novel and established chronic obstructive pulmonary disease loci. *Human molecular genetics*. 2015;24(4):1200-10. Epub 2014/10/16. doi: 10.1093/hmg/ddu525. PubMed PMID: 25315895.
 240. Glass K, **Quackenbush J**, Spentzos D, Haibe-Kains B, Yuan GC. A network model for angiogenesis in ovarian cancer. *BMC bioinformatics*. 2015;16:115. Epub 2015/04/19. doi: 10.1186/s12859-015-0551-y. PubMed PMID: 25888305; PubMed Central PMCID: PMC4408593.
 241. Lao T, Glass K, Qiu W, Polverino F, Gupta K, Morrow J, Mancini JD, Vuong L, Perrella MA, Hersh CP, Owen CA, **Quackenbush J**, Yuan GC, Silverman EK, Zhou X. Haploinsufficiency of Hedgehog interacting protein causes increased emphysema induced by cigarette smoke through network rewiring. *Genome medicine*. 2015;7(1):12. Epub 2015/03/13. doi: 10.1186/s13073-015-0137-3. PubMed PMID: 25763110; PubMed Central PMCID: PMC4355149.
 242. Olsen C, Fleming K, Prendergast N, Rubio R, Emmert-Streib F, Bontempi G, **Quackenbush J**, Haibe-Kains B. Using shRNA experiments to validate gene regulatory networks. *Genomics data*. 2015;4:123-6. doi: 10.1016/j.gdata.2015.03.011. PubMed PMID: 26484195; PubMed Central PMCID: PMC4535466.
 243. Padi M, **Quackenbush J**. Integrating transcriptional and protein interaction networks to prioritize condition-specific master regulators. *BMC systems biology*. 2015;9:80. doi: 10.1186/s12918-015-0228-1. PubMed PMID: 26576632; PubMed Central PMCID: PMC4650867.
 244. Berrios C, Padi M, Keibler MA, Park DE, Molla V, Cheng J, Lee SM, Stephanopoulos G, **Quackenbush J**, DeCaprio JA. Merkel Cell Polyomavirus Small T Antigen Promotes Pro-Glycolytic Metabolic Perturbations Required for Transformation. *PLoS Pathog*. 2016;12(11):e1006020. doi: 10.1371/journal.ppat.1006020. PubMed PMID: 27880818; PubMed Central PMCID: PMC4650867.
 245. Cloonan SM, Glass K, Lauchon-Contreras ME, Bhashyam AR, Cervo M, Pabon MA, Konrad C, Polverino F, Siempos II, Perez E, Mizumura K, Ghosh MC, Parameswaran H, Williams NC, Rooney KT, Chen ZH, Goldklang MP, Yuan GC, Moore SC, Demeo DL, Rouault TA, D'Armiento JM, Schon EA, Manfredi G, **Quackenbush J**, Mahmood A, Silverman EK, Owen CA, Choi AM. Mitochondrial iron chelation ameliorates cigarette smoke-induced bronchitis and emphysema in

- mice. *Nature medicine*. 2016;22(2):163-74. doi: 10.1038/nm.4021. PubMed PMID: 26752519; PubMed Central PMCID: PMC4742374.
246. De Rienzo A, Archer MA, Yeap BY, Dao N, Sciaranghella D, Sideris AC, Zheng Y, Holman AG, Wang YE, Dal Cin PS, Fletcher JA, Rubio R, Croft L, **Quackenbush J**, Sugarbaker PE, Munir KJ, Battilana JR, Gustafson CE, Chirieac LR, Ching SM, Wong J, Tay LC, Rudd S, Hercus R, Sugarbaker DJ, Richards WG, Bueno R. Gender-Specific Molecular and Clinical Features Underlie Malignant Pleural Mesothelioma. *Cancer research*. 2016;76(2):319-28. doi: 10.1158/0008-5472.CAN-15-0751. PubMed PMID: 26554828; PubMed Central PMCID: PMC4715909.
 247. Delaney SK, Hultner ML, Jacob HJ, Ledbetter DH, McCarthy JJ, Ball M, Beckman KB, Belmont JW, Bloss CS, Christman MF, Cosgrove A, Damiani SA, Danis T, Delledonne M, Dougherty MJ, Dudley JT, Faucett WA, Friedman JR, Haase DH, Hays TS, Heilsberg S, Huber J, Kaminsky L, Ledbetter N, Lee WH, Levin E, Libiger O, Linderman M, Love RL, Magnus DC, Martland A, McClure SL, Megill SE, Messier H, Nussbaum RL, Palaniappan L, Patay BA, Popovich BW, **Quackenbush J**, Savant MJ, Su MM, Terry SF, Tucker S, Wong WT, Green RC. Toward clinical genomics in everyday medicine: perspectives and recommendations. *Expert review of molecular diagnostics*. 2016;16(5):521-32. doi: 10.1586/14737159.2016.1146593. PubMed PMID: 26810587; PubMed Central PMCID: PMC4841021.
 248. Elias KM, Emori MM, Westerling T, Long H, Budina-Kolomets A, Li F, MacDuffie E, Davis MR, Holman A, Lawney B, Freedman ML, **Quackenbush J**, Brown M, Drapkin R. Epigenetic remodeling regulates transcriptional changes between ovarian cancer and benign precursors. *JCI Insight*. 2016;1(13). doi: 10.1172/jci.insight.87988. PubMed PMID: 27617304; PubMed Central PMCID: PMC4715158.
 249. Kusko RL, Brothers JF, 2nd, Tedrow J, Pandit K, Huleihel L, Perdomo C, Liu G, Juan-Guardela B, Kass D, Zhang S, Lenburg M, Martinez F, **Quackenbush J**, Sciurba F, Limper A, Geraci M, Yang I, Schwartz DA, Beane J, Spira A, Kaminski N. Integrated Genomics Reveals Convergent Transcriptomic Networks Underlying Chronic Obstructive Pulmonary Disease and Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine*. 2016;194(8):948-60. doi: 10.1164/rccm.201510-2026OC. PubMed PMID: 27104832; PubMed Central PMCID: PMC4715158.
 250. Malleshaiah M, Padi M, Rue P, **Quackenbush J**, Martinez-Arias A, Gunawardena J. Nac1 Coordinates a Sub-network of Pluripotency Factors to Regulate Embryonic Stem Cell Differentiation. *Cell reports*. 2016;14(5):1181-94. doi: 10.1016/j.celrep.2015.12.101. PubMed PMID: 26832399; PubMed Central PMCID: PMC4749452.
 251. Manimaran S, Selby HM, Okrah K, Ruberman C, Leek JT, **Quackenbush J**, Haibe-Kains B, Bravo HC, Johnson WE. BatchQC: interactive software for evaluating sample and batch effects in genomic data. *Bioinformatics*. 2016;32(24):3836-8. doi: 10.1093/bioinformatics/btw538. PubMed PMID: 27540268; PubMed Central PMCID: PMC4715158.
 252. Morrow JD, Cho MH, Hersh CP, Pinto-Plata V, Celli B, Marchetti N, Criner G, Bueno R, Washko G, Glass K, Choi AM, **Quackenbush J**, Silverman EK, DeMeo DL. DNA methylation profiling in human lung tissue identifies genes associated

- with COPD. *Epigenetics*. 2016;1-10. doi: 10.1080/15592294.2016.1226451. PubMed PMID: 27564456; PubMed Central PMCID: PMC5094634.
253. Platig J, Castaldi PJ, DeMeo D, **Quackenbush J**. Bipartite Community Structure of eQTLs. *PLoS computational biology*. 2016;12(9):e1005033. doi: 10.1371/journal.pcbi.1005033. PubMed PMID: 27618581; PubMed Central PMCID: PMC5019382.
 254. Safikhani Z, El-Hachem N, Quevedo R, Smirnov P, Goldenberg A, Juul Birkbak N, Mason C, Hatzis C, Shi L, Aerts HJ, **Quackenbush J**, Haibe-Kains B. Assessment of pharmacogenomic agreement. *F1000Res*. 2016;5:825. doi: 10.12688/f1000research.8705.1. PubMed PMID: 27408686; PubMed Central PMCID: PMC504926729.
 255. Safikhani Z, El-Hachem N, Smirnov P, Freeman M, Goldenberg A, Birkbak NJ, Beck AH, Aerts HJ, **Quackenbush J**, Haibe-Kains B. Safikhani et al. reply. *Nature*. 2016;540(7631):E2-E4. doi: 10.1038/nature19839. PubMed PMID: 27905430.
 256. Safikhani Z, El-Hachem N, Smirnov P, Freeman M, Goldenberg A, Birkbak NJ, Beck AH, Aerts HJ, **Quackenbush J**, Haibe-Kains B. Safikhani et al. reply. *Nature*. 2016;540(7631):E11-E2. doi: 10.1038/nature20581. PubMed PMID: 27905423.
 257. Safikhani Z, El-Hachem N, Smirnov P, Freeman M, Goldenberg A, Birkbak NJ, Beck AH, Aerts HJ, **Quackenbush J**, Haibe-Kains B. Safikhani et al. reply. *Nature*. 2016;540(7631):E6-E8. doi: 10.1038/nature20172. PubMed PMID: 27905416.
 258. van IDG, Glass K, **Quackenbush J**, Kuijjer ML. PyPanda: a Python package for gene regulatory network reconstruction. *Bioinformatics*. 2016;32(21):3363-5. doi: 10.1093/bioinformatics/btw422. PubMed PMID: 27402905; PubMed Central PMCID: PMC5079480.
 259. Vargas AJ, **Quackenbush J**, Glass K. Diet-induced weight loss leads to a switch in gene regulatory network control in the rectal mucosa. *Genomics*. 2016;108(3-4):126-33. doi: 10.1016/j.ygeno.2016.08.001. PubMed PMID: 27524493; PubMed Central PMCID: PMC5121035.
 260. Wu W, Parmar C, Grossmann P, **Quackenbush J**, Lambin P, Bussink J, Mak R, Aerts HJ. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. *Frontiers in oncology*. 2016;6:71. doi: 10.3389/fonc.2016.00071. PubMed PMID: 27064691; PubMed Central PMCID: PMC504811956.
 261. Campbell PT, Rebbeck TR, Nishihara R, Beck AH, Begg CB, Bogdanov AA, Cao Y, Coleman HG, Freeman GJ, Heng YJ, Huttenhower C, Irizarry RA, Kip NS, Michor F, Nevo D, Peters U, Phipps AI, Poole EM, Qian ZR, **Quackenbush J**, Robins H, Rogan PK, Slattery ML, Smith-Warner SA, Song M, VanderWeele TJ, Xia D, Zabor EC, Zhang X, Wang M, Ogino S. Proceedings of the third international molecular pathological epidemiology (MPE) meeting. *Cancer causes & control : CCC*. 2017;28(2):167-76. doi: 10.1007/s10552-016-0845-z. PubMed PMID: 28097472; PubMed Central PMCID: PMC5303153; <https://www.ncbi.nlm.nih.gov/pubmed/28097472>.
 262. Domenyuk V, Zhong Z, Stark A, Xiao N, O'Neill HA, Wei X, Wang J, Tinder TT, Tonapi S, Duncan J, Hornung T, Hunter A, Miglarese MR, Schorr J, Halbert DD, **Quackenbush J**, Poste G, Berry DA, Mayer G, Famulok M, Spetzler D. Plasma Exosome Profiling of Cancer Patients by a Next Generation Systems Biology Approach. *Sci Rep*. 2017;7:42741. doi: 10.1038/srep42741. PubMed PMID: 28218293; PubMed Central PMCID: PMC5316983; <https://www.ncbi.nlm.nih.gov/pubmed/28218293>.

263. Fagny M, Paulson JN, Kuijjer ML, Sonawane AR, Chen CY, Lopes-Ramos CM, Glass K, **Quackenbush J**, Platig J. Exploring regulation in tissues with eQTL networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2017;114(37):E7841-E50. doi: 10.1073/pnas.1707375114. PubMed PMID: 28851834; PubMed Central PMCID: PMC5604022; <https://www.ncbi.nlm.nih.gov/pubmed/28851834>.
264. Hill KE, Kelly AD, Kuijjer ML, Barry W, Rattani A, Garbutt CC, Kissick H, Janeway K, Perez-Atayde A, Goldsmith J, Gebhardt MC, Arredouani MS, Cote G, Hornicek F, Choy E, Duan Z, **Quackenbush J**, Haibe-Kains B, Spentzos D. An imprinted non-coding genomic cluster at 14q32 defines clinically relevant molecular subtypes in osteosarcoma across multiple independent datasets. *J Hematol Oncol*. 2017;10(1):107. doi: 10.1186/s13045-017-0465-4. PubMed PMID: 28506242; PubMed Central PMCID: PMC5433149; <https://www.ncbi.nlm.nih.gov/pubmed/28506242>.
265. Kibbe W, Klemm J, **Quackenbush J**. Cancer Informatics: New Tools for a Data-Driven Age in Cancer Research. *Cancer research*. 2017;77(21):e1-e2. doi: 10.1158/0008-5472.CAN-17-2212. PubMed PMID: 29092926; <https://www.ncbi.nlm.nih.gov/pubmed/29092926>.
266. Lopes-Ramos CM, Paulson JN, Chen CY, Kuijjer ML, Fagny M, Platig J, Sonawane AR, DeMeo DL, **Quackenbush J**, Glass K. Regulatory network changes between cell lines and their tissues of origin. *BMC genomics*. 2017;18(1):723. doi: 10.1186/s12864-017-4111-x. PubMed PMID: 28899340; PubMed Central PMCID: PMC5596945; <https://www.ncbi.nlm.nih.gov/pubmed/28899340>.
267. Morrow JD, Zhou X, Lao T, Jiang Z, DeMeo DL, Cho MH, Qiu W, Cloonan S, Pinto-Plata V, Celli B, Marchetti N, Criner GJ, Bueno R, Washko GR, Glass K, **Quackenbush J**, Choi AM, Silverman EK, Hersh CP. Functional interactors of three genome-wide association study genes are differentially expressed in severe chronic obstructive pulmonary disease lung tissue. *Sci Rep*. 2017;7:44232. doi: 10.1038/srep44232. PubMed PMID: 28287180; PubMed Central PMCID: PMC5347019; <https://www.ncbi.nlm.nih.gov/pubmed/28287180>.
268. Nishihara R, Glass K, Mima K, Hamada T, Nowak JA, Qian ZR, Kraft P, Giovannucci EL, Fuchs CS, Chan AT, **Quackenbush J**, Ogino S, Onnela JP. Biomarker correlation network in colorectal carcinoma by tumor anatomic location. *BMC bioinformatics*. 2017;18(1):304. doi: 10.1186/s12859-017-1718-5. PubMed PMID: 28623901; PubMed Central PMCID: PMC5474023; <https://www.ncbi.nlm.nih.gov/pubmed/28623901>.
269. Paulson JN, Chen CY, Lopes-Ramos CM, Kuijjer ML, Platig J, Sonawane AR, Fagny M, Glass K, **Quackenbush J**. Tissue-aware RNA-Seq processing and normalization for heterogeneous and sparse data. *BMC bioinformatics*. 2017;18(1):437. doi: 10.1186/s12859-017-1847-x. PubMed PMID: 28974199; PubMed Central PMCID: PMC5627434; <https://www.ncbi.nlm.nih.gov/pubmed/28974199>.
270. Qiu W, Guo F, Glass K, Yuan GC, **Quackenbush J**, Zhou X, Tantisira KG. Differential connectivity of gene regulatory networks distinguishes corticosteroid response in asthma. *J Allergy Clin Immunol*. 2017. doi: 10.1016/j.jaci.2017.05.052. PubMed PMID: 28736268; <https://www.ncbi.nlm.nih.gov/pubmed/28736268>.
271. Quiroz-Zarate A, Harshfield BJ, Hu R, Knoblauch N, Beck AH, Hankinson SE, Carey V, Tamimi RM, Hunter DJ, **Quackenbush J**, Hazra A. Expression

- Quantitative Trait loci (QTL) in tumor adjacent normal breast tissue and breast tumor tissue. *PloS one*. 2017;12(2):e0170181. doi: 10.1371/journal.pone.0170181. PubMed PMID: 28152060; PubMed Central PMCID: PMC5289428; <https://www.ncbi.nlm.nih.gov/pubmed/28152060>.
272. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, Ye Z, Makrigiorgos M, Fennessy F, Mak RH, Gillies R, **Quackenbush J**, Aerts H. Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer. *Cancer research*. 2017;77(14):3922-30. doi: 10.1158/0008-5472.CAN-17-0122. PubMed PMID: 28566328; PubMed Central PMCID: PMC5528160; <https://www.ncbi.nlm.nih.gov/pubmed/28566328>.
 273. Schlauch D, Paulson JN, Young A, Glass K, **Quackenbush J**. Estimating gene regulatory networks with pandaR. *Bioinformatics*. 2017;33(14):2232-4. doi: 10.1093/bioinformatics/btx139. PubMed PMID: 28334344; <https://www.ncbi.nlm.nih.gov/pubmed/28334344>.
 274. Sonawane AR, Platig J, Fagny M, Chen CY, Paulson JN, Lopes-Ramos CM, DeMeo DL, **Quackenbush J**, Glass K, Kuijjer ML. Understanding Tissue-Specific Gene Regulation. *Cell reports*. 2017;21(4):1077-88. doi: 10.1016/j.celrep.2017.10.001. PubMed PMID: 29069589; <https://www.ncbi.nlm.nih.gov/pubmed/29069589>.
 275. Wang YE, Kutnetsov L, Partensky A, Farid J, **Quackenbush J**. WebMeV: A Cloud Platform for Analyzing and Visualizing Cancer Genomic Data. *Cancer research*. 2017;77(21):e11-e4. doi: 10.1158/0008-5472.CAN-17-0802. PubMed PMID: 29092929; <https://www.ncbi.nlm.nih.gov/pubmed/29092929>.
 276. Yuan GC, Cai L, Elowitz M, Enver T, Fan G, Guo G, Irizarry R, Kharchenko P, Kim J, Orkin S, **Quackenbush J**, Saadatpour A, Schroeder T, Shivdasani R, Tirosch I. Challenges and emerging directions in single-cell analysis. *Genome biology*. 2017;18(1):84. doi: 10.1186/s13059-017-1218-y. PubMed PMID: 28482897; PubMed Central PMCID: PMC5421338; <https://www.ncbi.nlm.nih.gov/pubmed/28482897>.
 277. Barry JD, Fagny M, Paulson JN, Aerts H, Platig J, **Quackenbush J**. Histopathological Image QTL Discovery of Immune Infiltration Variants. *iScience*. 2018;5:80-9. Epub 2018/09/22. doi: 10.1016/j.isci.2018.07.001. PubMed PMID: 30240647; PubMed Central PMCID: PMC6123851.
 278. Carette X, Platig J, Young DC, Helmelt M, Young AT, Wang Z, Potluri LP, Moody CS, Zeng J, Prisc S, Paulson JN, Muntel J, Madduri AVR, Velarde J, Mayfield JA, Locher C, Wang T, **Quackenbush J**, Rhee KY, Moody DB, Steen H, Husson RN. Multisystem Analysis of Mycobacterium tuberculosis Reveals Kinase-Dependent Remodeling of the Pathogen-Environment Interface. *MBio*. 2018;9(2). Epub 2018/03/08. doi: 10.1128/mBio.02333-17. PubMed PMID: 29511081; PubMed Central PMCID: PMC5845002.
 279. Domenyuk V, Gatalica Z, Santhanam R, Wei X, Stark A, Kennedy P, Toussaint B, Levenberg S, Wang J, Xiao N, Greil R, Rinnerthaler G, Gampenrieder SP, Heimberger AB, Berry DA, Barker A, **Quackenbush J**, Marshall JL, Poste G, Vacirca JL, Vidal GA, Schwartzberg LS, Halbert DD, Voss A, Magee D, Miglarese MR, Famulok M, Mayer G, Spetzler D. Poly-ligand profiling differentiates trastuzumab-treated breast cancer patients according to their outcomes. *Nature*

- communications. 2018;9(1):1219. Epub 2018/03/25. doi: 10.1038/s41467-018-03631-z. PubMed PMID: 29572535; PubMed Central PMCID: PMC5865185.
280. Driesbaugh KH, Branchetti E, Grau JB, Keeney SJ, Glass K, Oyama MA, Rioux N, Ayoub S, Sacks MS, **Quackenbush J**, Levy RJ, Ferrari G. Serotonin receptor 2B signaling with interstitial cell activation and leaflet remodeling in degenerative mitral regurgitation. *J Mol Cell Cardiol.* 2018;115:94-103. Epub 2018/01/02. doi: 10.1016/j.yjmcc.2017.12.014. PubMed PMID: 29291394; PubMed Central PMCID: PMC5856457.
 281. Hicks SC, Okrah K, Paulson JN, **Quackenbush J**, Irizarry RA, Bravo HC. Smooth quantile normalization. *Biostatistics.* 2018;19(2):185-98. Epub 2017/10/17. doi: 10.1093/biostatistics/kxx028. PubMed PMID: 29036413; PubMed Central PMCID: PMC5862355.
 282. Hosny A, Parmar C, **Quackenbush J**, Schwartz LH, Aerts H. Artificial intelligence in radiology. *Nat Rev Cancer.* 2018;18(8):500-10. Epub 2018/05/20. doi: 10.1038/s41568-018-0016-5. PubMed PMID: 29777175; PubMed Central PMCID: PMC6268174.
 283. Kuijjer ML, Paulson JN, Salzman P, Ding W, **Quackenbush J**. Cancer subtype identification using somatic mutation data. *Br J Cancer.* 2018;118(11):1492-501. Epub 2018/05/17. doi: 10.1038/s41416-018-0109-7. PubMed PMID: 29765148; PubMed Central PMCID: PMC5988673.
 284. Lopes-Ramos CM, Kuijjer ML, Ogino S, Fuchs CS, DeMeo DL, Glass K, **Quackenbush J**. Gene Regulatory Network Analysis Identifies Sex-Linked Differences in Colon Cancer Drug Metabolism. *Cancer Res.* 2018;78(19):5538-47. Epub 2018/10/03. doi: 10.1158/0008-5472.CAN-18-0454. PubMed PMID: 30275053; PubMed Central PMCID: PMC6169995.
 285. Morrow JD, Cho MH, Platig J, Zhou X, DeMeo DL, Qiu W, Celli B, Marchetti N, Criner GJ, Bueno R, Washko GR, Glass K, **Quackenbush J**, Silverman EK, Hersh CP. Ensemble genomic analysis in human lung tissue identifies novel genes for chronic obstructive pulmonary disease. *Hum Genomics.* 2018;12(1):1. Epub 2018/01/18. doi: 10.1186/s40246-018-0132-z. PubMed PMID: 29335020; PubMed Central PMCID: PMC5769240.
 286. Morrow JD, Glass K, Cho MH, Hersh CP, Pinto-Plata V, Celli B, Marchetti N, Criner G, Bueno R, Washko G, Choi AMK, **Quackenbush J**, Silverman EK, DeMeo DL. Human Lung DNA Methylation Quantitative Trait Loci Colocalize with Chronic Obstructive Pulmonary Disease Genome-Wide Association Loci. *Am J Respir Crit Care Med.* 2018;197(10):1275-84. Epub 2018/01/10. doi: 10.1164/rccm.201707-1434OC. PubMed PMID: 29313708; PubMed Central PMCID: PMC5955059.
 287. Padi M, **Quackenbush J**. Detecting phenotype-driven transitions in regulatory network structure. *NPJ Syst Biol Appl.* 2018;4:16. Epub 2018/05/01. doi: 10.1038/s41540-018-0052-5. PubMed PMID: 29707235; PubMed Central PMCID: PMC5908977.
 288. Parmar C, Barry JD, Hosny A, **Quackenbush J**, Aerts H. Data Analysis Strategies in Medical Imaging. *Clinical cancer research : an official journal of the American*

- Association for Cancer Research. 2018;24(15):3492-9. Epub 2018/03/28. doi: 10.1158/1078-0432.CCR-18-0385. PubMed PMID: 29581134; PubMed Central PMCID: PMC6082690.
289. Qiu W, Guo F, Glass K, Yuan GC, **Quackenbush J**, Zhou X, Tantisira KG. Differential connectivity of gene regulatory networks distinguishes corticosteroid response in asthma. *J Allergy Clin Immunol*. 2018;141(4):1250-8. Epub 2017/07/25. doi: 10.1016/j.jaci.2017.05.052. PubMed PMID: 28736268; PubMed Central PMCID: PMC5775928.
 290. Campbell PT, Ambrosone CB, Nishihara R, Aerts H, Bondy M, Chatterjee N, Garcia-Closas M, Giannakis M, Golden JA, Heng YJ, Kip NS, Koshiol J, Liu XS, Lopes-Ramos CM, Mucci LA, Nowak JA, Phipps AI, **Quackenbush J**, Schoen RE, Sholl LM, Tamimi RM, Wang M, Weijenberg MP, Wu CJ, Wu K, Yao S, Yu KH, Zhang X, Rebbeck TR, Ogino S. Proceedings of the fourth international molecular pathological epidemiology (MPE) meeting. *Cancer Causes Control*. 2019;30(8):799-811. Epub 2019/05/10. doi: 10.1007/s10552-019-01177-z. PubMed PMID: 31069578; PubMed Central PMCID: PMC6614001.
 291. Gaynor SM, Sun R, Lin X, **Quackenbush J**. Identification of differentially expressed gene sets using the Generalized Berk-Jones statistic. *Bioinformatics*. 2019;35(22):4568-76. Epub 2019/05/08. doi: 10.1093/bioinformatics/btz277. PubMed PMID: 31062858; PubMed Central PMCID: PMC6853697.
 292. Hersh CP, Adcock IM, Celedon JC, Cho MH, Christiani DC, Himes BE, Kaminski N, Mathias RA, Meyers DA, **Quackenbush J**, Redline S, Steiling KA, Tabor HK, Tobin MD, Wurfel MM, Yang IV, Koppelman GH. High-Throughput Sequencing in Respiratory, Critical Care, and Sleep Medicine Research. An Official American Thoracic Society Workshop Report. *Ann Am Thorac Soc*. 2019;16(1):1-16. Epub 2018/12/29. doi: 10.1513/AnnalsATS.201810-716WS. PubMed PMID: 30592451; PubMed Central PMCID: PMC6812157.
 293. Kuijjer ML, Hsieh PH, **Quackenbush J**, Glass K. lionessR: single sample network inference in R. *BMC Cancer*. 2019;19(1):1003. Epub 2019/10/28. doi: 10.1186/s12885-019-6235-7. PubMed PMID: 31653243; PubMed Central PMCID: PMC6815019.
 294. Kuijjer ML, Tung MG, Yuan G, **Quackenbush J**, Glass K. Estimating Sample-Specific Regulatory Networks. *iScience*. 2019;14:226-40. Epub 2019/04/15. doi: 10.1016/j.isci.2019.03.021. PubMed PMID: 30981959; PubMed Central PMCID: PMC6463816.
 295. Lozano Calderon SA, Garbutt C, Kim J, Lietz CE, Chen YL, Bernstein K, Chebib I, Nielsen GP, Deshpande V, Rubio R, Wang YE, **Quackenbush J**, Delaney T, Raskin K, Schwab J, Cote G, Spentzos D. Clinical and Molecular Analysis of Pathologic Fracture-associated Osteosarcoma: MicroRNA profile Is Different and Correlates with Prognosis. *Clin Orthop Relat Res*. 2019;477(9):2114-26. Epub 2019/08/08. doi: 10.1097/CORR.0000000000000867. PubMed PMID: 31389890; PubMed Central PMCID: PMC7000091.
 296. Schwede M, Waldron L, Mok SC, Wei W, Basunia A, Merritt MA, Mitsiades CS, Parmigiani G, Harrington DP, **Quackenbush J**, Birrer MJ, Culhane AC. The

- impact of stroma admixture on molecular subtypes and prognostic gene signatures in serous ovarian cancer. *Cancer Epidemiol Biomarkers Prev.* 2019. Epub 2019/12/25. doi: 10.1158/1055-9965.EPI-18-1359. PubMed PMID: 31871106.
297. St Hilaire MA, Kristal BS, Rahman SA, Sullivan JP, **Quackenbush J**, Duffy JF, Barger LK, Gooley JJ, Czeisler CA, Lockley SW. Using a Single Daytime Performance Test to Identify Most Individuals at High-Risk for Performance Impairment during Extended Wake. *Sci Rep.* 2019;9(1):16681. Epub 2019/11/15. doi: 10.1038/s41598-019-52930-y. PubMed PMID: 31723161; PubMed Central PMCID: PMC6853981.
 298. Altenbuchinger M, Weihs A, **Quackenbush J**, Grabe HJ, Zacharias HU. Gaussian and Mixed Graphical Models as (multi-)omics data analysis tools. *Biochim Biophys Acta Gene Regul Mech.* 2020;1863(6):194418. Epub 2019/10/23. doi: 10.1016/j.bbagr.2019.194418. PubMed PMID: 31639475; PubMed Central PMCID: PMC7166149.
 299. Fagny M, Platig J, Kuijjer ML, Lin X, **Quackenbush J**. Nongenetic cancer-risk SNPs affect oncogenes, tumour-suppressor genes, and immune function. *Br J Cancer.* 2020;122(4):569-77. Epub 2019/12/07. doi: 10.1038/s41416-019-0614-3. PubMed PMID: 31806877; PubMed Central PMCID: PMC7028992.
 300. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Massive Analysis Quality Control Society Board of D, Waldron L, Wang B, McIntosh C, Goldenberg A, Kundaje A, Greene CS, Broderick T, Hoffman MM, Leek JT, Korthauer K, Huber W, Brazma A, Pineau J, Tibshirani R, Hastie T, Ioannidis JPA, **Quackenbush J**, Aerts H. Transparency and reproducibility in artificial intelligence. *Nature.* 2020;586(7829):E14-E6. Epub 2020/10/16. doi: 10.1038/s41586-020-2766-y. PubMed PMID: 33057217.
 301. Kuijjer ML, Fagny M, Marin A, **Quackenbush J**, Glass K. PUMA: PANDA Using MicroRNA Associations. *Bioinformatics.* 2020;36(18):4765-73. Epub 2020/08/30. doi: 10.1093/bioinformatics/btaa571. PubMed PMID: 32860050; PubMed Central PMCID: PMC7750953.
 302. Lam HC, Cloonan SM, Bhashyam AR, Haspel JA, Singh A, Sathirapongsasuti JF, Cervo M, Yao H, Chung AL, Mizumura K, An CH, Shan B, Franks JM, Haley KJ, Owen CA, Tesfaigzi Y, Washko GR, **Quackenbush J**, Silverman EK, Rahman I, Kim HP, Mahmood A, Biswal SS, Ryter SW, Choi AM. Histone deacetylase 6-mediated selective autophagy regulates COPD-associated cilia dysfunction. *The Journal of clinical investigation.* 2020;130(11):6189. Epub 2020/11/03. doi: 10.1172/JCI143863. PubMed PMID: 33136096; PubMed Central PMCID: PMC7598037.
 303. Lietz CE, Garbutt C, Barry WT, Deshpande V, Chen YL, Lozano-Calderon SA, Wang Y, Lawney B, Ebb D, Cote GM, Duan Z, Hornicek FJ, Choy E, Petur Nielsen G, Haibe-Kains B, **Quackenbush J**, Spentzos D. MicroRNA-mRNA networks define translatable molecular outcome phenotypes in osteosarcoma. *Sci Rep.* 2020;10(1):4409. Epub 2020/03/12. doi: 10.1038/s41598-020-61236-3. PubMed PMID: 32157112; PubMed Central PMCID: PMC7064533.

304. Lopes-Ramos CM, Chen CY, Kuijjer ML, Paulson JN, Sonawane AR, Fagny M, Platig J, Glass K, **Quackenbush J**, DeMeo DL. Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. *Cell Rep*. 2020;31(12):107795. Epub 2020/06/25. doi: 10.1016/j.celrep.2020.107795. PubMed PMID: 32579922.
305. Lopes-Ramos CM, **Quackenbush J**, DeMeo DL. Genome-Wide Sex and Gender Differences in Cancer. *Front Oncol*. 2020;10:597788. Epub 2020/12/18. doi: 10.3389/fonc.2020.597788. PubMed PMID: 33330090; PubMed Central PMCID: PMC7719817.
306. Morrow JD, Make B, Regan E, Han M, Hersh CP, Tal-Singer R, **Quackenbush J**, Choi AMK, Silverman EK, DeMeo DL. DNA Methylation Is Predictive of Mortality in Current and Former Smokers. *Am J Respir Crit Care Med*. 2020;201(9):1099-109. Epub 2020/01/30. doi: 10.1164/rccm.201902-0439OC. PubMed PMID: 31995399; PubMed Central PMCID: PMC7193844.
307. Paulson JN, Williams BL, Hehnlly C, Mishra N, Sinnar SA, Zhang L, Ssentongo P, Mbabazi-Kabachelor E, Wijetunge DSS, von Bredow B, Mulondo R, Kiwanuka J, Bajunirwe F, Bazira J, Bebell LM, Burgoine K, Couto-Rodriguez M, Ericson JE, Erickson T, Ferrari M, Gladstone M, Guo C, Haran M, Hornig M, Isaacs AM, Kaaya BN, Kangere SM, Kulkarni AV, Kumbakumba E, Li X, Limbrick DD, Jr., Magombe J, Morton SU, Mugamba J, Ng J, Olupot-Olupot P, Onen J, Peterson MR, Roy F, Sheldon K, Townsend R, Weeks AD, Whalen AJ, **Quackenbush J**, Ssenyonga P, Galperin MY, Almeida M, Atkins H, Warf BC, Lipkin WI, Broach JR, Schiff SJ. Paenibacillus infection with frequent viral coinfection contributes to postinfectious hydrocephalus in Ugandan infants. *Sci Transl Med*. 2020;12(563). Epub 2020/10/02. doi: 10.1126/scitranslmed.aba0565. PubMed PMID: 32998967; PubMed Central PMCID: PMC7774825.
308. Schwede M, Waldron L, Mok SC, Wei W, Basunia A, Merritt MA, Mitsiades CS, Parmigiani G, Harrington DP, **Quackenbush J**, Birrer MJ, Culhane AC. The Impact of Stroma Admixture on Molecular Subtypes and Prognostic Gene Signatures in Serous Ovarian Cancer. *Cancer Epidemiol Biomarkers Prev*. 2020;29(2):509-19. Epub 2019/12/25. doi: 10.1158/1055-9965.EPI-18-1359. PubMed PMID: 31871106; PubMed Central PMCID: PMC7448721.
309. Silverman EK, Schmidt H, Anastasiadou E, Altucci L, Angelini M, Badimon L, Balligand JL, Benincasa G, Capasso G, Conte F, Di Costanzo A, Farina L, Fiscon G, Gatto L, Gentili M, Loscalzo J, Marchese C, Napoli C, Paci P, Petti M, **Quackenbush J**, Tieri P, Viggiano D, Vilahur G, Glass K, Baumbach J. Molecular networks in Network Medicine: Development and applications. *Wiley Interdiscip Rev Syst Biol Med*. 2020;12(6):e1489. Epub 2020/04/21. doi: 10.1002/wsbm.1489. PubMed PMID: 32307915.
310. Burkholz R, **Quackenbush J**, Bojar D. Using graph convolutional neural networks to learn a representation for glycans. *Cell Rep*. 2021;35(11):109251. Epub 2021/06/17. doi: 10.1016/j.celrep.2021.109251. PubMed PMID: 34133929.
311. Hatchi E, Goehring L, Landini S, Skourti-Stathaki K, DeConti DK, Abderazzaq FO, Banerjee P, Demers TM, Wang YE, **Quackenbush J**, Livingston DM. BRCA1 and

- RNAi factors promote repair mediated by small RNAs and PALB2-RAD52. *Nature*. 2021;591(7851):665-70. Epub 2021/02/05. doi: 10.1038/s41586-020-03150-2. PubMed PMID: 33536619; PubMed Central PMCID: PMC8245199.
312. Lopes-Ramos CM, Belova T, Brunner TH, Ben Guebila M, Osorio D, **Quackenbush J**, Kuijjer ML. Regulatory Network of PD1 Signaling Is Associated with Prognosis in Glioblastoma Multiforme. *Cancer Res*. 2021;81(21):5401-12. Epub 2021/09/09. doi: 10.1158/0008-5472.CAN-21-0730. PubMed PMID: 34493595; PubMed Central PMCID: PMC8563450.
 313. Neri de Souza Reis V, Tahira AC, Daguano Gastaldi V, Mari P, Portolese J, Feio Dos Santos AC, Lisboa B, Mari J, Caetano SC, Brunoni D, Bordini D, Silvestre de Paula C, Vencio RZN, **Quackenbush J**, Brentani H. Environmental Influences Measured by Epigenetic Clock and Vulnerability Components at Birth Impact Clinical ASD Heterogeneity. *Genes (Basel)*. 2021;12(9). Epub 2021/09/29. doi: 10.3390/genes12091433. PubMed PMID: 34573415; PubMed Central PMCID: PMC8467464.
 314. Snyder JM, Huang RY, Bai H, Rao VR, Cornes S, Barnholtz-Sloan JS, Gutman D, Fasano R, Van Meir EG, Brat D, Eschbacher J, **Quackenbush J**, Wen PY, Lee JW. Analysis of morphological characteristics of IDH-mutant/wildtype brain tumors using whole-lesion phenotype analysis. *Neurooncol Adv*. 2021;3(1):vdab088. Epub 2021/08/20. doi: 10.1093/noajnl/vdab088. PubMed PMID: 34409295; PubMed Central PMCID: PMC8367280.
 315. Sonawane AR, DeMeo DL, **Quackenbush J**, Glass K. Constructing gene regulatory networks using epigenetic data. *NPJ Syst Biol Appl*. 2021;7(1):45. Epub 2021/12/11. doi: 10.1038/s41540-021-00208-3. PubMed PMID: 34887443; PubMed Central PMCID: PMC8660777.
 316. Weighill D, Ben Guebila M, Glass K, Platig J, Yeh JJ, **Quackenbush J**. Gene Targeting in Disease Networks. *Front Genet*. 2021;12:649942. Epub 2021/05/11. doi: 10.3389/fgene.2021.649942. PubMed PMID: 33968133; PubMed Central PMCID: PMC8103030.
 317. Weighill D, Guebila MB, Lopes-Ramos C, Glass K, **Quackenbush J**, Platig J, Burkholz R. Gene regulatory network inference as relaxed graph matching. *Proc Conf AAAI Artif Intell*. 2021;35(11):10263-72. Epub 2021/10/29. PubMed PMID: 34707916; PubMed Central PMCID: PMC8546743.
 318. Young AT, Carette X, Helmelt M, Steen H, Husson RN, **Quackenbush J**, Platig J. Multi-omic regulatory networks capture downstream effects of kinase inhibition in *Mycobacterium tuberculosis*. *NPJ Syst Biol Appl*. 2021;7(1):8. Epub 2021/01/31. doi: 10.1038/s41540-020-00164-4. PubMed PMID: 33514755; PubMed Central PMCID: PMC7846781.
 319. Ben Guebila M, Lopes-Ramos CM, Weighill D, Sonawane AR, Burkholz R, Shamsaei B, Platig J, Glass K, Kuijjer ML, **Quackenbush J**. GRAND: a database of gene regulatory network models across human conditions. *Nucleic Acids Res*. 2022;50(D1):D610-D21. Epub 2021/09/12. doi: 10.1093/nar/gkab778. PubMed PMID: 34508353; PubMed Central PMCID: PMC8728257.

320. Ben Guebila M, Weighill D, Lopes-Ramos CM, Burkholz R, Pop RT, Palepu K, Shapoval M, Fagny M, Schlauch D, Glass K, Altenbuchinger M, Kuijjer ML, Platig J, **Quackenbush J**. An online notebook resource for reproducible inference, analysis and publication of gene regulatory networks. *Nat Methods*. 2022;19(5):511-3. Epub 2022/04/24. doi: 10.1038/s41592-022-01479-2. PubMed PMID: 35459940; PubMed Central PMCID: PMC9239854.
321. Gaynor SM, Fagny M, Lin X, Platig J, **Quackenbush J**. Connectivity in eQTL networks dictates reproducibility and genomic properties. *Cell Rep Methods*. 2022;2(5):100218. Epub 2022/06/01. doi: 10.1016/j.crmeth.2022.100218. PubMed PMID: 35637906; PubMed Central PMCID: PMC9142682.
322. Guebila MB, Morgan DC, Glass K, Kuijjer ML, DeMeo DL, **Quackenbush J**. gpuZoo: Cost-effective estimation of gene regulatory networks using the Graphics Processing Unit. *NAR Genom Bioinform*. 2022;4(1):lqac002. Epub 2022/02/15. doi: 10.1093/nargab/lqac002. PubMed PMID: 35156023; PubMed Central PMCID: PMC8826808.
323. Moll M, Hobbs BD, Menon A, Ghosh AJ, Putman RK, Hino T, Hata A, Silverman EK, **Quackenbush J**, Castaldi PJ, Hersh CP, McGeachie MJ, Sin DD, Tal-Singer R, Nishino M, Hatabu H, Hunninghake GM, Cho MH. Blood gene expression risk profiles and interstitial lung abnormalities: COPDGene and ECLIPSE cohort studies. *Respir Res*. 2022;23(1):157. Epub 2022/06/18. doi: 10.1186/s12931-022-02077-8. PubMed PMID: 35715807; PubMed Central PMCID: PMC9204872.
324. Weighill D, Ben Guebila M, Glass K, **Quackenbush J**, Platig J. Predicting genotype-specific gene regulatory networks. *Genome Res*. 2022;32(3):524-33. Epub 2022/02/24. doi: 10.1101/gr.275107.120. PubMed PMID: 35193937; PubMed Central PMCID: PMC8896459.
325. Yao S, Campbell PT, Ugai T, Gierach G, Abubakar M, Adalsteinsson V, Almeida J, Brennan P, Chanock S, Golub T, Hanash S, Harris C, Hathaway CA, Kelsey K, Landi MT, Mahmood F, Newton C, **Quackenbush J**, Rodig S, Schultz N, Tearney G, Tworoger SS, Wang M, Zhang X, Garcia-Closas M, Rebbeck TR, Ambrosone CB, Ogino S. Proceedings of the fifth international Molecular Pathological Epidemiology (MPE) meeting. *Cancer Causes Control*. 2022;33(8):1107-20. Epub 2022/06/28. doi: 10.1007/s10552-022-01594-7. PubMed PMID: 35759080; PubMed Central PMCID: PMC9244289.
326. Shutta KH, Weighill D, Burkholz R, Guebila MB, DeMeo DL, Zacharias HU, **Quackenbush J**, Altenbuchinger M. DRAGON: Determining Regulatory Associations using Graphical models on multi-Omic Networks. *Nucleic Acids Res*. 2023;51(3):e15. Epub 2022/12/20. doi: 10.1093/nar/gkac1157. PubMed PMID: 36533448; PubMed Central PMCID: PMC9943674. Shutta KH, Weighill D, Burkholz R, Guebila MB, DeMeo DL, Zacharias HU, Quackenbush J, Altenbuchinger M. DRAGON: Determining Regulatory Associations using Graphical models on multi-Omic Networks. *Nucleic Acids Res*. 2023;51(3):e15. Epub 2022/12/20. doi: 10.1093/nar/gkac1157. PubMed PMID: 36533448; PubMed Central PMCID: PMC9943674.

327. Saha E, Guebila MB, Fanfani V, Fischer J, Shutta KH, Mandros P, DeMeo DL, **Quackenbush J**, Lopes-Ramos CM. Gene regulatory Networks Reveal Sex Difference in Lung Adenocarcinoma. *bioRxiv*. 2023. Epub 20230924. doi: 10.1101/2023.09.22.559001. PubMed PMID: 37790409; PubMed Central PMCID: PMC10543009.
328. Hossain I, Fanfani V, **Quackenbush J**, Burkholz R. Biologically informed NeuralODEs for genome-wide regulatory dynamics. *bioRxiv*. 2023. Epub 2023/03/14. doi: 10.1101/2023.02.24.529835. PubMed PMID: 36909563; PubMed Central PMCID: PMC10002636.
329. Hossain I, Fanfani V, **Quackenbush J**, Burkholz R. Biologically informed NeuralODEs for genome-wide regulatory dynamics. *Res Sq*. 2023. Epub 2023/03/31. doi: 10.21203/rs.3.rs-2675584/v1. PubMed PMID: 36993392; PubMed Central PMCID: PMC10055646 authors declare no competing interests.
330. Ben Guebila M, Wang T, Lopes-Ramos CM, Fanfani V, Weighill D, Burkholz R, Schlauch D, Paulson JN, Altenbuchinger M, Shutta KH, Sonawane AR, Lim J, Calderer G, van IDGP, Morgan D, Marin A, Chen CY, Song Q, Saha E, DeMeo DL, Padi M, Platig J, Kuijjer ML, Glass K, **Quackenbush J**. The Network Zoo: a multilingual package for the inference and analysis of gene regulatory networks. *Genome Biol*. 2023;24(1):45. Epub 2023/03/10. doi: 10.1186/s13059-023-02877-1. PubMed PMID: 36894939; PubMed Central PMCID: PMC9999668.

Preprints:

1. Glass K, **Quackenbush J**, Kepner J. High Performance Computing of Gene Regulatory Networks using a Message-Passing Model. 2015.
2. Kuijjer M, Tung M, Yuan G, **Quackenbush J**, Glass K. Estimating sample-specific regulatory networks. eprint arXiv:150506440. 2015. doi: citeulike-article-id:13659997. <https://arxiv.org/abs/1505.06440>.
3. Safikhani Z, Freeman M, Smirnov P, El-Hachem N, She A, Quevedo R, Goldenberg A, Juul Birkbak N, Hatzis C, Shi L, Beck AH, Aerts HJ, **Quackenbush J**, Haibe-Kains B. Revisiting inconsistency in large pharmacogenomic studies. *bioRxiv*. 2015. doi: 10.1101/026153. <http://biorxiv.org/content/early/2015/09/11/026153>.
4. Chen C-Y, Lopes-Ramos CM, Kuijjer ML, Paulson JN, Sonawane AR, Fagny M, Platig J, Glass K, **Quackenbush J**, DeMeo DL. Sexual dimorphism in gene expression and regulatory networks across human tissues. *bioRxiv*. 2016. doi: 10.1101/082289. <http://biorxiv.org/content/early/2016/10/20/082289>.
5. Fagny M, Paulson JN, Kuijjer ML, Sonawane AR, Chen C-Y, Lopes-Ramos CM, Glass K, **Quackenbush J**, Platig J. A network-based approach to eQTL interpretation and SNP functional characterization. *bioRxiv*. 2016. doi: 10.1101/086587. <http://biorxiv.org/content/early/2016/11/09/086587>.
6. Hicks SC, Okrah K, Paulson JN, **Quackenbush J**, Irizarry RA, Corrada Bravo H. Smooth Quantile Normalization. *bioRxiv*. 2016. doi: 10.1101/085175. <http://biorxiv.org/content/early/2016/11/03/085175>.

7. Lopes-Ramos CM, Paulson JN, Chen C-Y, Kuijjer ML, Fagny M, Platig J, Sonawane AR, DeMeo DL, **Quackenbush J**, Glass K. Transcriptional landscape of cell lines and their tissues of origin. *bioRxiv*. 2016. doi: 10.1101/082065. <http://biorxiv.org/content/early/2016/10/20/082065>.
8. Safikhani Z, El-Hachem N, Quevedo R, Smirnov P, Goldenberg A, Juul Birkbak N, Mason CE, Hatzis C, Shi L, Aerts HJ, **Quackenbush J**, Haibe-Kains B. Assessment of pharmacogenomic agreement. *bioRxiv*. 2016. doi: 10.1101/048470. <http://biorxiv.org/content/early/2016/04/13/048470>.
9. Schlauch D, Glass K, Hersh CP, Silverman EK, **Quackenbush J**. Estimating Drivers of Cell State Transitions Using Gene Regulatory Network Models. *bioRxiv*. 2016. doi: 10.1101/089003. <http://biorxiv.org/content/early/2016/11/21/089003>.
10. Barry JD, Fagny M, Paulson JN, Aerts H, Platig J, **Quackenbush J**. Histopathological image QTL discovery of thyroid autoimmune disease variants. *bioRxiv*. 2017. doi: 10.1101/126730. <https://www.biorxiv.org/content/biorxiv/early/2017/08/08/126730.full.pdf>.
11. Kuijjer ML, Paulson JN, Salzman P, Ding W, **Quackenbush J**. Cancer subtype identification using somatic mutation data. *bioRxiv*. 2017. doi: 10.1101/228031. <https://www.biorxiv.org/content/biorxiv/early/2017/12/03/228031.full.pdf>.
12. Padi M, **Quackenbush J**. Phenotype-Driven Transitions In Regulatory Network Structure. *bioRxiv*. 2017. doi: 10.1101/142281. <http://www.biorxiv.org/content/early/2017/05/25/142281>.
13. Qiang J, Ding W, **Quackenbush J**, Chen P. Network-based Distance Metric with Application to Discover Disease Subtypes in Cancer. 2017.
14. Sonawane AR, Platig J, Fagny M, Chen C-Y, Paulson JN, Lopes-Ramos CM, DeMeo DL, **Quackenbush J**, Glass K, Kuijjer ML. Understanding Tissue-specific Gene Regulation. *bioRxiv*. 2017. doi: 10.1101/110601. <http://biorxiv.org/content/early/2017/02/21/110601>.
15. Wang YE, Kuznetsov L, Partensky A, Farid J, **Quackenbush J**. WebMeV: A Cloud Platform for Analyzing and Visualizing Cancer Genomic Data. *bioRxiv*. 2017. doi: 10.1101/147884.
16. Yuan G-C, Cai L, Elowitz M, Enver T, Fan G, Guo G, Irizarry R, Kharchenko P, Kim J, Orkin S, **Quackenbush J**, Saadatpour A, Schroeder T, Shivdasani R, Tirosch I. Challenges And Emerging Directions In Single-Cell Analysis. *bioRxiv*. 2017. doi: 10.1101/127761. <https://www.biorxiv.org/content/biorxiv/early/2017/04/15/127761.full.pdf>.
17. Fagny M, Platig J, Kuijjer ML, Lin X, **Quackenbush J**. Nongenetic cancer-risk SNPs affect oncogenes, tumor suppressor genes, and immune function. *bioRxiv*. 2018.
18. Lopes-Ramos CM, Kuijjer ML, Ogino S, Fuchs CS, DeMeo DL, Glass K, **Quackenbush J**. Gene regulatory network analysis identifies sex-linked differences in colon cancer drug metabolism processes. *bioRxiv*. 2018. doi: 10.1101/277186. <https://www.biorxiv.org/content/biorxiv/early/2018/03/07/277186.full.pdf>.
19. Schwede M, Waldron L, Mok SC, Wei W, Basunia A, Merritt MA, Parmigiani G, Harrington D, **Quackenbush J**, Birrer MJ, Culhane A. The impact of stroma on the

- discovery of molecular subtypes and prognostic gene signatures in serous ovarian cancer. *bioRxiv*. 2018.
20. Gaynor S, Fagny M, Lin X, Platig J, **Quackenbush J**. Connectivity of variants in eQTL networks dictates reproducibility and functionality. *bioRxiv*. 2019.
 21. Kuijjer ML, **Quackenbush J**, Glass K. lionessR: single-sample network reconstruction in R. *bioRxiv*. 2019:582098. doi: 10.1101/582098.
 22. Young AT, Carette X, Helmelt M, Steen H, Husson RN, **Quackenbush J**, Platig J. Multi-omic regulatory networks capture downstream effects of kinase inhibition in *Mycobacterium tuberculosis*. *bioRxiv*. 2019:584177. doi: 10.1101/584177.
 23. Guebila MB, Glass K, Kuijjer ML, Platig J, Lopes-Ramos C, Padi M, Fagny M, Sonawane AR, Paulson J, Burkholz R, Weighill D, Altenbuchinger M, Wang T, **Quackenbush J**. The Network Zoo: Integrated Tools for Network Inference and Analysis 2020. Available from: <https://netzoo.github.io/>.
 24. Guebila MB, Glass K, Kuijjer ML, Platig J, Lopes-Ramos C, Padi M, Fagny M, Sonawane AR, Paulson J, Burkholz R, Weighill D, Altenbuchinger M, Wang T, **Quackenbush J**. netBooks: A Cloud Resource for Training in Network Inference and Analysis 2020. Available from: <http://netbooks.networkmedicine.org>.
 25. Guebila MB, Lopes-Ramos C, Burkholz R, Weighill D, Wang T, **Quackenbush J**. GRAND: The Gene Regulatory Network Database 2020. Available from: <https://grand.networkmedicine.org/>.
 26. Sonawane AR, DeMeo DL, **Quackenbush J**, Glass K. Constructing Gene Regulatory Networks using Epigenetic Data. *bioRxiv*. 2020:2020.10.19.345827. doi: 10.1101/2020.10.19.345827.
 27. Wang YE, Kutnetsov L, Partensky A, Farid J, **Quackenbush J**. WebMeV software for large-scale genomic data analysis 2020. Available from: <http://mev.tm4.org>.
 28. Weighill D, Guebila MB, Lopes-Ramos C, Glass K, **Quackenbush J**, Platig J, Burkholz R. Gene Regulatory Network Inference as Relaxed Graph Matching. *bioRxiv*. 2020:2020.06.23.167999. doi: 10.1101/2020.06.23.167999.
 29. Burkholz R, **Quackenbush J**, Bojar D. Using Graph Convolutional Neural Networks to Learn a Representation for Glycans. *bioRxiv*. 2021:2021.03.01.433491. doi: 10.1101/2021.03.01.433491.
 30. Lopes-Ramos C, Belova T, Brunner T, **Quackenbush J**, Kuijjer ML. Regulation of PD1 signaling is associated with prognosis in glioblastoma multiforme. *bioRxiv*. 2021:2021.02.11.430786. doi: 10.1101/2021.02.11.430786.
 31. Weighill D, Guebila MB, Glass K, **Quackenbush J**, Platig J. Predicting genotype-specific gene regulatory networks. *bioRxiv*. 2021:2021.01.18.427134. doi: 10.1101/2021.01.18.427134.
 32. Weighill DA, Burkholz R, Guebila MB, Zacharias HU, **Quackenbush J**, Altenbuchinger M. DRAGON: Determining Regulatory Associations using Graphical models on multi-Omic Networks. *arXiv*. 2021;arXiv:2104.01690.
 33. Weighill DA, Guebila MB, Glass K, Platig J, Yeh J, **Quackenbush J**. Gene targeting in disease networks. *arXiv*. 2021;arXiv:2101.03985.
 34. Lück N, Lohmayer R, Solbrig S, Schrod S, Wipfler T, Shutta KH, Guebila MB, Schäfer A, Reißbarth T, Zacharias HU, Oefner PJ, **Quackenbush J**, Altenbuchinger

- M. SpaCeNet: Spatial Cellular Networks from omics data. bioRxiv. 2022:2022.09.01.506219. doi: 10.1101/2022.09.01.506219.
35. Guebila MB, Wang T, Lopes-Ramos CM, Fanfani V, Weighill D, Burkholz R, Schlauch D, Paulson JN, Altenbuchinger M, Sonanwane A, Lim J, Calderer G, van Ijzendoorn D, Morgan D, Marin A, Chen C-Y, Song A, Shutta K, DeMeo D, Padi M, Platig J, Kuijjer ML, Glass K, **Quackenbush J**. The Network Zoo: a multilingual package for the inference and analysis of biological networks. bioRxiv. 2022:2022.05.30.494077. doi: 10.1101/2022.05.30.494077.

Reviews:

1. **Quackenbush J**. Viva la revolution! A report from the FANTOM meeting. *Nature Genet.* 2000;26(3):255-6.
2. **Quackenbush J**. Computational analysis of microarray data. *Nat Rev Genet.* 2001;2(6):418-27.
3. **Quackenbush J**. Microarray data normalization and transformation. *Nat Genet.* 2002;32 Suppl:496-501
4. Dudoit S, Gentleman RC, **Quackenbush J**. Open source software for the analysis of microarray data. *Biotechniques.* 2003;Suppl:45-51.
5. **Quackenbush J**. Extracting meaning from functional genomics experiments. *Toxicol Appl Pharmacol.* 2005;207(2 Suppl):195-9.
6. **Quackenbush J**. Microarray analysis and tumor classification. *N Engl J Med.* 2006;354(23):2463-72.
7. **Quackenbush J**. From 'omes to biology. *Anim Genet.* 2006;37 Suppl 1:48-56.
8. **Quackenbush J**. Computational approaches to analysis of DNA microarray data. *Methods Inf Med.* 2006;45 Suppl 1:91-103.
9. **Quackenbush J**. Extracting biology from high-dimensional biological data. *J Exp Biol.* 2007;210(Pt 9):1507-17.
10. Bateman, A., and **Quackenbush, J.** (2009). Bioinformatics for next generation sequencing. *Bioinformatics* 25, 429.

Editorials:

1. **Quackenbush J**. Data standards for 'omic' science. *Nat Biotechnol.* 2004;22(5):613-4.
2. **Quackenbush J**, Salzberg SL. It is time to end the patenting of software. *Bioinformatics.* 2006;22(12):1416-7.
3. **Quackenbush J**. Weighing our measures of gene expression. *Mol Syst Biol.* 2006;2:63.
4. **Quackenbush J**. Standardizing the standards. *Mol Syst Biol.* 2006;2:2006 0010.
5. **Quackenbush J**, Irizarry RA. Response to Shields: 'MIAME, we have a problem'. *Trends Genet.* 2006;22(9):471-2.
6. **Quackenbush J**. Learning to share. *Scientific American.* 2014;311(1):S22. Epub 2014/07/01. PubMed PMID: 24974706.

7. **Quackenbush J.** Perspective: Learning to share. *Nature*. 2014;509(7502):S68. Epub 2014/05/30. doi: 10.1038/509S68a. PubMed PMID: 24870827; PubMed Central PMCID: PMC4205237.
8. Kolovos C, Poznansky M, **Quackenbush J**, Lal N. Docs offer playbook for getting kids back in school, *Boston Herald*, January 15, 2021.

Books:

1. Coulston H. C., **Quackenbush J.**, Brazma A., **Microarray Gene Expression Data Analysis: A Beginner's Guide**, Blackwell Publishing, 2003.
2. Kelley, J. M. and **Quackenbush J.**, "Sequencing PCR Products," in **PCR Methods Manual**, Second edition, M. Innis, D. Gelfand, and J. Sninsky, eds., Academic Press, 1999.
3. **Quackenbush J.**, Adams M. D., Fraser C. M., and Venter J. C., "Genome Projects: From Microbes to Man, From Sequence to Function," in **Genomics: Commercial Opportunities from a Scientific Revolution**, G. K. Dixon, L. G. Copping, and D. Livingstone, eds. BIOS Scientific Publishers, 1998.
4. **Quackenbush, J.** Using DNA Microarrays to Assay Gene Expression. In: A. Baxevanis and B. Ouellette (eds.), *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, third edition: John Wiley and Sons, 2004.
5. **Quackenbush, J.** The Human Genome: Book of Essential Knowledge. Imagine! Publishing, 2011.

Educational Material: I have taught numerous courses and workshops and make the materials freely available. I developed and co-taught a HarvardX course on reproducible research (<https://www.edx.org/course/principles-statistical-computational-harvardx-ph527x>).

Thesis: Gauge Field Theory in Two Space-Time Dimensions: Anomalies and Applications to String Models.

Nonprint Materials, i.e., film strips, films, videotapes and computer-based materials relevant to appointee's academic field. *Include description of item, who commissioned, purpose, users, penetration in field.*

None

Patents

None

Abstracts (*only those containing data not yet published in complete form.*)

None

Cases in which I have testified at deposition or trial.

To the best of my knowledge, this is a complete list. The parties with whom I was engaged as an expert are shown in bold.

1. **Foresight Diagnostics Inc.** v. *Personalis Inc.*, Petitions for Post-Grant Review Case No. IPR2023-00224
2. **Foresight Diagnostics Inc.** v. *Personalis Inc.*, Petitions for Post-Grant Review Case No. IPR2023-00317
3. *Twinstrand Biosciences* v. **Guardant Health**, IPR2022-01400
4. *RegenxBio Inc. and The Trustees of the University of Pennsylvania* v. **Sarepta Therapeutics, Inc. and Sarepta Therapeutics Three, LLC.**, Case No. C.A. No. 20-1226 (RGA)
5. **Natera, Inc.** v. *CareDx, Inc.*, Case No. C.A. No. 20-038 (CFC) (CJB) CONSOLIDATED
6. **10x Genomics, Inc. and Prognosys. Biosciences, Inc.**, v. *Nanostring Technologies, Inc.* Civil Action No. 21-cv-653-MFK
7. **Guardant Health, Inc.** v. *Twinstrand Biosciences, Inc. and the University of Washington*, Case No. IPR2022-01158
8. **Guardant Health, Inc.** v. *Twinstrand Biosciences, Inc. and the University of Washington*, Case No. IPR2022-00817
9. **Guardant Health, Inc.** v. *University of Washington Through its Center for Commercialization*, Petitions for Post-Grant Review Case No. IPR2022-00450
10. *Twinstrand Biosciences, Inc., and the University of Washington* v. **Guardant Health, Inc.** Case No. C.A. No. 21-1126-VAC-SRF
11. *Natera, Inc.* v. *ArcherDX, Inc.*, Case No. 1:20-CV-125 LPS and *Natera, Inc.* v. *Genosity, Inc.*, Case No. 20-CV-1352 LPS
12. **BioSpyder Technologies, Inc.** v. *HTG Molecular Diagnostics, Inc.*, Civil Action No. CV-21-34-TUC-RCC
13. *Illumina, Verinata Health, and Stanford University* v. **Ariosa Diagnostics and Roche Molecular Systems**, Case No. 3:12-cv-05501-SI (consolidated with Case No. 3:14-cv-01921-SI and Case No. 3:15-cv-02216-SI)
14. *RainDance Technologies Inc.* v. **10x Genomics, Inc.**, (D. Del. Case No. 15-cv-00152)
15. *Bio-Rad Laboratories and the University of Chicago* v. **10x Genomics**, C.A. No. 15-152-RGA (Markman Hearing and at Trial)
16. *Juno Therapeutics and Memorial Sloan Kettering* v. **Kite Pharma**, Case No. 2:17-cv-07639-SJO-RAO
17. *Foundation Medicine* v. **Guardant Health**, Civil Action No. 2:16-cv-00523
18. *Foundation Medicine* v. **Guardant Health**, Case No. IPR2019-00652
19. *Foundation Medicine* v. **Guardant Health**, Case No. IPR2019-00634
20. **Guardant Health, Inc.** v. *Foundation Medicine, Inc.*, Case Nos. IPR2017-01170, IPR2017-01447, IPR2017-01448
21. *Illumina* v. **Natera**, Case No. IPR2019-01201
22. *CareDx, Inc. et al* v. **Eurofins Viracor, Inc.**, D. Del., C.A. No.: 1:19-cv-1804-CFC
23. *CareDx, Inc., et al.* v. **Natera, Inc.**, Case No. 1:19-cv-00567-CFC (D.Del.) and *CareDx, Inc.* v. **Natera, Inc.**, Case No. 1:19-cv-00622-CFC (D. Del.)

24. **Guardant Health, Inc.** v. *University of Washington Through its Center for Commercialization*, Petitions for Post-Grant Review/ *Guardant Health v. Keygene N.V.*, Case No. IPR2020-00674
25. **10x Genomics, Inc.** v. *Celsee, Inc.*, Case No. 1:19-cv-00862-CFC
26. *Ravgen, Inc. v. Natera, Inc. and NSTX, Inc.*, C.A. No. AU:20-cv-00692-ADA
27. *Bio-Rad Laboratories v. 10x Genomics*, Case No.1:19-cv-12533-WGY
28. *Bio-Rad Laboratories Inc. et al. v. 10x Geomics, Inc.*, C.A. No. 15-152-RGA (D. Del.)
29. *Progenity, Inc v. Natera, Inc.*, Case No. IPR2021-00267
30. *Applera Corporation v. Illumina, Inc. et al.*, Case No. 3:07-cv-02845 (N.D. Cal.)
31. *Affymetrix Inc. v. Illumina Inc.*, Case No. 1:04-cv-00901-JJF (D. Del.)
32. *Affymetrix, Inc. v. HYSEQ, Inc.*, Case No. 5:99-cv-21163-JF (N.D. Cal.)
33. **HYSEQ, Inc.** v. *Affymetrix, Inc.*, Case No. 5:00-cv-20050-JF (N.D. Cal.)

Cases in the past five years for which I have submitted sworn declarations in the past five years, but not testified at deposition or trial.

To the best of my recollection, this is a complete list. The parties with whom I was engaged as an expert are shown in bold.

1. **Foresight Diagnostics Inc.** v. *Personalis Inc.*, Petitions for Post-Grant Review Case No. IPR2023-00545
2. **Foresight Diagnostics Inc.** v. *Personalis Inc.*, Petitions for Post-Grant Review Case No. IPR2023-00546
3. *Illumina, Inc., et al. v. BGI Genomics Co., Ltd., et al.*, Case Nos. 3:19-cv-03770-WHO and 3:20-cv-01465-WHO (N.D. Cal.)
4. **10x Genomics** - Inter partes review of U.S. Patent No. 10,190,115, Docket No.: IPR2021-00133