厚生労働科学研究委託費

革新的がん医療実用化研究事業

（委託業務題目）クリニカルシークエンスのための実用的なバイオインフォマティクスプログラムの開発および情報解析に関する研究

平成26年度　委託業務成果報告書

業務主任者　加藤　護

平成27（2015）年　3月

本報告書は、厚生労働省の厚生労働科学研究委託事業による委託業務として、独立行政法人国立がん研究センターが実施した平成２６年度「クリニカルシークエンスのための実用的なバイオインフォマティクスプログラムの開発および情報解析に関する研究」の成果を取りまとめたものです。

目　　次

別紙３

厚生労働科学研究委託費（革新的がん医療実用化研究事業）
委託業務成果報告（総括）

クリニカルシークエンスのための実用的なバイオインフォマティクスプログラムの開発および情報解析に関する研究

研究要旨

本研究では、臨床で使われる低腫瘍率かつFFPE 用の変異検出アルゴリズムを開発し、国立がん研究センターの臨床シークエンスに適用し、その結果をもって、日本人のがん変異に関する特徴を調べることである。

業務項目の担当責任者氏名・所属研究機関名及び所属研究機関における職名

国立がん研究センター研究所
部門長　加藤護

国立がん研究センター研究所
分野長　河野隆志

A．研究目標

近年、国際がんゲノム・コンソーシアムといった大規模プロジェクトによって、様々ながんにおける体細胞変異が網羅的に同定されつつある。国際がんゲノム・コンソーシアムにおいては申請者も情報解析に参加し、次世代シークエンサーを使って肝がんの体細胞変異とその性質を解明しつつある(Totoki et al, 2011, Nature Genetics; Fujimoto et al, 2012, Nature Genetics; Totoki et al, 2014, Nature Genetics)。

これらの成果を臨床に応用する方法として、現在臨床シークエンスが実現されつつある。臨床シークエンスとは、患者のがん組織試料を次世代シークエンサーによって配列決定し、検出されたがん変異に対する作用薬の知識と突き合わせて、より適切な治療法や臨床試験のよりよい層別化を行うことである。アメリカのベンチャー企業Foundation Medicine 社は既に商用運用を開始し、研究開発の成果をNature Biotechnology 誌(Frampton et al, 2013)に報告している。国立がん研究センターでは、2013年度より臨床シークエンスの実施基盤を独自に開発し、その運用が開始されつつある。

臨床シークエンスは、次世代シークエンサーによる各個人の配列決定であり膨大なデータを生む。膨大なデータを処理するにはコンピュータによる情報解析が不可欠であるが、研究用に開発されたアルゴリズムを臨床シークエンスにそのまま適用は出来ない。臨床用の情報解析には、研究用の情報解析では現れなかった問題が出てくるからである。申請者はこれまで、当センターの臨床シークエンス用に点変異・融合遺伝子検出のアルゴリズムを開発してきたが、これは研究用で使われる高腫瘍率、かつ、質の良い凍結試料をベースに開発されたものである。臨床で使われる低腫瘍率、かつ、質の悪いFFPE 試料での検討はまだ不十分である。また、コピー数異常に関しては、FFPE 試料に対する検出のアルゴリズムさえ定まっていない。さらに、アルゴリズムを適用した結果を整理して、日本人の診断に役立つ情報を抽出する分析も不十分である。

本研究の目的は、臨床で使われる低腫瘍率かつFFPE 用の点変異・融合遺伝子・コピー数異常検出のアルゴリズム群を開発し（平成26-27年度）、臨床シークエンスに適用して日本人がん変異のデータベースを作成し（平成27-28年度）、日本人の変異に関する情報解析を行うこと（H28年度）である。

B．研究方法

研究協力者である国立がん研究センター病院の山本昇医師、田村研治医師がFFPE の患者臨床試料を取得し、同病院の担当病理医が腫瘍率推定を含む病理組織の評価を行う。分担研究者である河野隆志博士、研究協力者である市川仁博士が試料のDNAをキャプチャし、次世代シークエンサーによって配列決定する。2週間当たり約5名の患者の試料が、この流れでシークエンスされた。

配列決定されたデータ（fastq）を基に、低腫瘍率かつFFPE 用のアルゴリズム開発およびその情報解析を行う。

厚生労働科学研究委託費（革新的がん医療実用化研究事業）
委託業務成果報告（総括）

本研究は国立がんセンター倫理審査委員会にてすでに承認が得られている。試料採取に関しては、病理診断・検査の残余を研究に用いるため、提供者に新たに侵襲を与えず、また診断への影響や治療への介入はない。臨床試料の提供者には、試料が医学研究に使われることを文書および口頭で説明し、個別に同意を取得する。臨床試料は連結可能匿名化を行い、個人が特定されないようにして解析に用いる。

Ｃ．研究成果

・FFPE 試料の収集およびシークエンス

国立がん研究センター病院の医師の協力の下、サンプルが収集され、病理医による確認、DNA 抽出と品質チェックを経て、約100 サンプルが次世代シークエンサーによって、シークエンスされた。

・点突然変異（SNV/indel）

研究用の場合、凍結サンプルというコストは高いが質は良いサンプルを使う。しかし臨床シークエンスの場合、低コストだがDNA が変性してしまうFFPE サンプルを使う。FFPE サンプルを使うとデータにエラーが多く、ミスアライメントを引き起こしやすい。また、臨床応用においては低腫瘍率のサンプルも使わざるを得ない。腫瘍率が低いサンプルでは当然感度が落ちる。

これらの問題に対処するために、今回ミスアライメントを特異的に除外するフィルターなど、11種類のフィルターを開発した。これらのフィルターにおいては、様々な種類のデータに対し頑健に対応するため、統計的手法が最大限利用されている。

この検出アルゴリズムを、上述したサンプルに適用し、変異を検出した。IGV やさまざな統計数値を調べ、アルゴリズムの改良を行った。次に、検出された変異を質量分析法（MassArray）によって検証した。その結果、SNV/indel 共に、非常な高精度で検出されていることを確認した。

・融合遺伝子検出

融合遺伝子検出では、これまで開発したsingle-end sequence read ベースのアルゴリズムに、pair-end sequence read ベースのアルゴリズムを組み合わせ、感度を高める改良を行っている。

・コピー数変化検出

COSMIC とDGV データベースを調べ、既知のがんでコピー数変動がない領域、日本人において生殖細胞系列でコピー数変動がない領域を選び出し、この領域をコピー数の基準領域とするアルゴリズムを開発した。また、データからコピー数変化領域を自動的に決定するsegmentation アルゴリズムを開発している。

Ｄ．考察

FFPE 試料のシークエンスのさい、中には解析できないほどDNA の質が悪いものがあることが分かった。これらは現在、DNA の質を実際に測定することで解決している。点突然変異検出アルゴリズムは、様々な改良と確認実験を経て、実用レベルに達したと言える。

Ｅ．結論

本年度の研究計画通り、FFPE試料のDNAを実際に約100 検体ほどシークエンスし、点変異検出アルゴリズムを実用化した。しかし今後も精度改良に努めていく予定である。今後は、融合遺伝子検出、および、コピー数変異検出アルゴリズムの改良に重点を移していく。これらの確認実験も行う。

Ｆ．健康危険情報

特になし

１．論文発表

（次のページに続く）

別紙３

厚生労働科学研究委託費（革新的がん医療実用化研究事業）
委託業務成果報告（総括）

Ｇ．研究発表

Yasushi Totoki, Kenji Tatsuno, Kyle R Covington, Hiroki Ueda, Chad J Creighton, Mamoru Kato, Shingo Tsuji, Lawrence A Donehower, Betty L Slagle, Hiromi Nakamura, Shogo Yamamoto, Eve Shinbrot, Natsuko Hama, Megan Lehmkuhl, Fumie Hosoda, Yasuhito Arai, Kim Walker, Mahmoud Dahdouli, Kengo Gotoh, Genta Nagae, Marie-Claude Gingras, Donna M Muzny, Hidenori Ojima, Kazuaki Shimada, Yutaka Midorikawa, John A Goss, Ronald Cotton, Akimasa Hayashi, Junji Shibahara, Shumpei Ishikawa, Jacfranz Guiteau, Mariko Tanaka, Tomoko Urushidate, Shoko Ohashi, Naoko Okada, Harsha Doddapaneni, Min Wang, Yiming Zhu, Huyen Dinh, Takuji Okusaka, Norihiro Kokudo, Tomoo Kosuge, Tadatoshi Takayama, Masashi Fukayama, Richard A Gibbs, David A Wheeler, Hiroyuki Aburatani & Tatsuhiro Shibata
Trans-ancestry mutational landscape of hepatocellular carcinoma genomes.
Nature Genetics, 2014, 46, 1267-1273.

Tatsuji Mizukami, Kouya Shiraishi, Yoko Shimada, Hideaki Ogiwara, Koji Tsuta, Hitoshi Ichikawa, Hiromi Sakamoto, Mamoru Kato, Tatsuhiro Shibata, Takashi Nakano, Takashi Kohno,

Molecular mechanisms underlying oncogenic RET fusion in lung adenocarcinoma.
Journal of Thoracic Oncology, 2014, 9, 622-630.

加藤 護. 総説：「一細胞ゲノム解析」医学の歩み、2014, 249, 1088-1092. (2014)

翻訳：加藤　護、「発がんドライバー変異の同定」 by David Tamborero, Abel Gonzalez-Perez and Nuria Lopez-Bigas, 実験医学、2014, 32, 213-219

Nakaoku T, Tsuta K, Ichikawa H, Shiraishi K, Sakamoto H, Enari M, Furuta K, Shimada Y, Ogiwara H, Watanabe SI, Nokihara H, Yasuda K, Hiramoto M, Nammo T, Ishigame T, Schetter AJ, Okayama H, Harris CC, Kim YH, Mishima M, Yokota J, Yoshida T, Kohno T* (2014) Druggable oncogene fusions in invasive mucinous lung adenocarcinoma. *Clin Cance Res.* 20(12) :3087- 3093.

Saito M, Ishigame T, Tsuta K, Kumamoto K, Imai T, Kohno T* (2014) A mouse model of KIF5B-RET fusion-dependent lung tumorigenesis. *Carcinogenesis.* 35(11):2452-2456

Saito M, Shiraishi K, Matsumoto K, Schetter AJ, Ogata-Kawata H, Tsuchiya N, Kunitoh H, Nokihara H, Watanabe SI, Tsuta K, Kumamoto K, Takenoshita S, Yokota J, Harris CC, Kohno T*(2014) A Three-microRNA signature predicts responses to platinum-Based doublet chemotherapy in patients with lung adenocarcinoma. Clin Cancer Res. 20(18); 4784- 93.

２．学会発表

○加藤 護. A bioinformatics system developed for cancer clinical sequencing in National Cancer Center, Japan European Society for Medical Oncology, 招待講演, Madrid, Spain, September 26-30, 2014.

○加藤 護. Development of a computational system in cancer clinical sequencing. 招待講演, 2014 Cancer Symposium, Mokpo, Korea, April 15-19, 2014.

○加藤　護. A bioinformatics system developed for clinical sequencing in National Cancer Center, Japan (Symposia)
73rd Annual Meeting of the Japanese Cancer Association, Yokohama, Japan, September 25-27, 2014.

厚生労働科学研究委託費（革新的がん医療実用化研究事業）
委託業務成果報告（総括）

○中奥敬史、蔦幸治、渡邉俊一、軒原浩、金永学、三嶋理晃、横田淳、河野隆志：Lung invasive mucinous adenocarcinoma(IMA)における治療標的となる新規遺伝子融合、第55回日本肺癌学会学術総会、京都、第55回日本肺癌学会学術総会抄録集、P116、11月、2014年

H．知的財産権の出願・登録状況
（予定を含む。）

別紙４

厚生労働科学研究委託費（革新的がん医療実用化研究事業）
委託業務成果報告（業務項目）

シークエンス情報解析

業務主任者又は担当責任者　加藤護　国立がん研究センター研究所部門長

研究要旨

本研究では、クリニカルシークエンスに適用できるような低腫瘍率・およびFFPE 試料に対する変異検出アルゴリズムの開発、および情報解析を行う。

A．研究目的

本研究では、臨床で使われる低腫瘍率またはFFPE 用の点変異・融合遺伝子・コピー数異常を検出するアルゴリズムを開発し、臨床検体のシークエンス・データに適用して正確な変異検出と情報解析を行う。

B．研究方法

国立がん研究センターで得られたFFPE臨床試料を、研究分担者が次世代シークエンサーによって配列決定する。このデータを用い、IGV や様々な統計数値を用いて異常を示すパターンやシグナルがあったばあい、フィルターの改良を行って、異常なパターンやシグナルを消していく。このように変異検出アルゴリズムを改良していく。最終的に、質量分析法（MassArray）によって確認実験を行う。

C．研究結果

シークエンスされたFFPE 約１００サンプルに対し、上の手続きを行って、点変異検出アルゴリズムの改良を行った。最終的にマスアレイで検証を行い、高精度に検出できることを確認した。このアルゴリズムは、実際の臨床シークエンスで使われた。

D．考察

今後、他ツールとの徹底的な比較を進めていく。また、融合遺伝子、コピー数異常検出アルゴリズムの改良、および検証実験も進めていく。

E．結論

FFPE サンプルに特化した点変異検出アルゴリズムを開発・改良し、これが点変異を高精度に検出できることを確認した。これにより、実際の臨床シークエンスで使用可能であると判断できた。

F．健康危険情報

特になし

G．研究発表

1.　論文発表

Yasushi Totoki, Kenji Tatsuno, Kyle R Covington, Hiroki Ueda, Chad J Creighton, Mamoru Kato, Shingo Tsuji, Lawrence A Donehower, Betty L Slagle, Hiromi Nakamura, Shogo Yamamoto, Eve Shinbrot, Natsuko Hama, Megan Lehmkuhl, Fumie Hosoda, Yasuhito Arai, Kim Walker, Mahmoud Dahdouli, Kengo Gotoh, Genta Nagae, Marie-Claude Gingras, Donna M Muzny, Hidenori Ojima, Kazuaki Shimada, Yutaka Midorikawa, John A Goss, Ronald Cotton, Akimasa Hayashi, Junji Shibahara, Shumpei Ishikawa, Jacfranz Guiteau, Mariko Tanaka, Tomoko Urushidate, Shoko Ohashi, Naoko Okada, Harsha Doddapaneni, Min Wang, Yiming

厚生労働科学研究委託費（革新的がん医療実用化研究事業）
委託業務成果報告（業務項目）

Zhu, Huyen Dinh, Takuji Okusaka, Norihiro Kokudo, Tomoo Kosuge, Tadatoshi Takayama, Masashi Fukayama, Richard A Gibbs, David A Wheeler, Hiroyuki Aburatani & Tatsuhiro Shibata
Trans-ancestry mutational landscape of hepatocellular carcinoma genomes.
Nature Genetics, 2014, 46, 1267-1273.

Tatsuji Mizukami, Kouya Shiraishi, Yoko Shimada, Hideaki Ogiwara, Koji Tsuta, Hitoshi Ichikawa, Hiromi Sakamoto, Mamoru Kato, Tatsuhiro Shibata, Takashi Nakano, Takashi Kohno,

Molecular mechanisms underlying oncogenic RET fusion in lung adenocarcinoma.
Journal of Thoracic Oncology, 2014, 9, 622-630.

加藤 護. 総説:「一細胞ゲノム解析」医学の歩み、2014, 249, 1088-1092. (2014)

翻訳：加藤　護、「発がんドライバー変異の同定」 by David Tamborero, Abel Gonzalez-Perez and Nuria Lopez-Bigas, 実験医学、2014, 32, 213-219

2.　学会発表

○A bioinformatics system developed for cancer clinical sequencing in National Cancer Center, Japan、uropean Society for Medical Oncology, Madrid, Spain, September 26-30, 2014.

○Development of a computational system in cancer clinical sequencing. 2014 Cancer Symposium, Mokpo, Korea, April 15-19, 2014.

○A bioinformatics system developed for clinical sequencing in National Cancer Center, Japan、横浜、第73回日本癌学会学術総会

H．知的財産権の出願・登録状況
（予定を含む。）

なし

別紙４

厚生労働科学研究委託費（革新的がん医療実用化研究事業）
委託業務成果報告（業務項目）

シークエンス情報解析における検証

業務主任者又は担当責任者　河野 隆志　国立がん研究センター研究所分野長

研究要旨

本研究では、クリニカル・シークエンスに必要な情報解析アルゴリズムを開発するための大本となる、臨床FFPE 試料からの次世代シークエンサー・データを取得し、さらに、アルゴリズムの精度を確認する実験を行う。

Ａ．研究目的

国立がん研究センターに構築された臨床シークエンスの実施基盤に基づき、FFPE の臨床検体からDNA 試料を取得し、次世代シークエンサーからシークエンスデータを得る。また、開発されたアルゴリズムの妥当性を計る確認実験を行う。

Ｂ．研究方法

臨床シークエンスにおいて、医師より患者臨床試料をFFPE組織として収集する。収集した組織からDNA が抽出され、次世代シーケンサにより、配列決定される。

配列データは、研究代表者が開発したアルゴリズムによって、変異が同定される。

点変異検出アルゴリズムの検証としては、質量分析法（MassArray）によって検証を行う。

融合遺伝子検出アルゴリズムの検証としては、PCR 実験によって融合遺伝子の確認を行う。

コピー数変化アルゴリズムの検証としては、qPCR によって増幅率を確認する。

（倫理面への配慮）

研究に利用する手術標本は、研究対象者から同意を得た上で、検体は匿名化し、患者に不利益がないよう、プライバシーを厳守して行う。臨床情報も同様に匿名化しプライバシーに配慮する。更に、研究成果は個人情報が公開されないように発表・報告する。

すべての研究は、「個人情報保護法」ならびに「ヒトゲノム・遺伝子解析研究に関する倫理指針」「臨床指針」「疫学指針」を尊守し、あらかじめ倫理委員会での承認手続きを行った上で進めている。

Ｃ．研究結果

医師より患者臨床試料を得、約100 のFFPE 臨床試料に対し、DNA を抽出し、quality を測定した後、シークエンスを行い、配列データを得た。研究代表者が開発したアルゴリズムに対し、マスアレイによる確認、PCR 実験による確認、qPCR による確認を行った。

Ｄ．考察

FFPE 組織からのDNA 試料の中にはシークエンスしても解析不可能なほど質が悪いものが存在していることが分かった。これは、DNA のquality をあらかじめ測ることにより、解決できた。

Ｅ．結論

国立がん研究センターに構築された臨床シークエンスの実施基盤に基づき、FFPE の臨床検体から、次世代シークエンサーによるシークエンス・データを取

厚生労働科学研究委託費（革新的がん医療実用化研究事業）
委託業務成果報告（業務項目）

得することに成功した。また、アルゴリズムを確認する実験も実施できた。

Ｆ．健康危険情報
特になし

Ｇ．研究発表

1.　論文発表

Nakaoku T, Tsuta K, Ichikawa H, Shiraishi K, Sakamoto H, Enari M, Furuta K, Shimada Y, Ogiwara H, Watanabe SI, Nokihara H, Yasuda K, Hiramoto M, Nammo T, Ishigame T, Schetter AJ, Okayama H, Harris CC, Kim YH, Mishima M, Yokota J, Yoshida T, Kohno T* (2014) Druggable oncogene fusions in invasive mucinous lung adenocarcinoma. *Clin Cance Res.* 20(12) :3087- 3093.

Saito M, Ishigame T, Tsuta K, Kumamoto K, Imai T, Kohno T* (2014) A mouse model of KIF5B-RET fusion-dependent lung tumorigenesis. *Carcinogenesis.* 35(11):2452-2456

Saito M, Shiraishi K, Matsumoto K, Schetter AJ, Ogata-Kawata H, Tsuchiya N, Kunitoh H, Nokihara H, Watanabe SI, Tsuta K, Kumamoto K, Takenoshita S, Yokota J, Harris CC, Kohno T*(2014) A Three-microRNA signature predicts responses to platinum-Based doublet chemotherapy in patients with lung adenocarcinoma. Clin Cancer Res. 20(18); 4784- 93.

2.　学会発表

○中奥敬史、蔦幸治、渡邉俊一、軒原浩、金永学、三嶋理晃、横田淳、河野隆志：Lung invasive mucinous adenocarcinoma (IMA)における治療標的となる新規遺伝子融合、第55回日本肺癌学会学術総会、京都、第55回日本肺癌学会学術総会抄録集、P116、11月、2014年

Ｈ．知的財産権の出願・登録状況
（予定を含む。）

1. 特許取得
なし

2. 実用新案登録
なし

3.その他
なし

学 会 等 発 表 実 績

委託業務題目「 クリニカルシークエンスのための実用的なバイオインフォマティクスプログラムの開発および情報解析」
機関名　独立行政法人 国立がん研究センター

１．学会等における口頭・ポスター発表

発表した成果（発表題目、口頭・ポスター発表の別）	発表者氏名	発表した場所（学会等名）	発表した時期	国内・外の別
A bioinformatics system developed for cancer clinical sequencing in National Cancer Center, Japan.	加藤護、黒坂功、柴田龍弘、土原一哉、田村研治、山本尚吾、市川仁、河野隆志	European Society for Medical Oncology	2014	国外
Development of a computational system in cancer clinical sequencing.	加藤護	Cancer symposium	2014	国外
A bioinformatics system developed for clinical sequencing in National Cancer Center, Japan.	加藤護、黒坂功、柴田龍弘、土原一哉、田村研治、山本尚吾、市川仁、河野隆志	第73回日本癌学会学術総会	2014	国内
Lung invasive mucinous adenocarcinoma(IMA)における治療標的となる新規遺伝子融合	中奥敬史、蔦幸治、渡邉俊一、軒原浩、金永学、三嶋理晃、横田淳、河野隆志	第55回日本肺癌学会学術総会	2014	国内

２．学会誌・雑誌等における論文掲載

掲載した論文（発表題目）	発表者氏名	発表した場所	発表した時期	国内・外の別
Trans-ancestry mutational landscape of hepatocellular carcinoma genomes.	Yasushi Totoki, Kenji Tatsuno, Kyle R Covington, Hiroki Ueda, Chad J Creighton, Mamoru Kato, Shingo Tsuji, Lawrence A Donehower, Betty L Slagle, Hiromi Nakamura, Shogo Yamamoto, Eve Shinbrot, Natsuko Hama, Megan Lehmkuhl, Fumie Hosoda, Yasuhito Arai, Kim Walker, Mahmoud Dahdouli, Kengo Gotoh, Genta Nagae, Marie-Claude Gingras, Donna M Muzny, Hidenori Ojima, Kazuaki Shimada, Yutaka Midorikawa, John A Goss, Ronald Cotton, Akimasa Hayashi, Junji Shibahara, Shumpei Ishikawa, Jacfranz Guiteau, Mariko Tanaka, Tomoko Urushidate, Shoko Ohashi, Naoko Okada, Harsha Doddapaneni, Min Wang, Yiming Zhu, Huyen Dinh, Takuji Okusaka, Norihiro Kokudo, Tomoo Kosuge, Tadatoshi Takayama, Masashi Fukayama, Richard A Gibbs, David A Wheeler, Hiroyuki Aburatani & Tatsuhiro Shibata	Nature Genetics	2014	国外

Molecular mechanisms underlying oncogenic RET fusion in lung adenocarcinoma.	Tatsuji Mizukami, Kouya Shiraishi, Yoko Shimada, Hideaki Ogiwara, Koji Tsuta, Hitoshi Ichikawa, Hiromi Sakamoto, Mamoru Kato, Tatsuhiro Shibata, Takashi Nakano, Takashi Kohno,	Journal of Thoracic Oncology	2014	国外
一細胞ゲノム解析	加藤護	医学の歩み	2014	国内
発がんドライバー変異の同定	翻訳：加藤　護（原著者：David Tamborero, Abel Gonzalez-Perez and Nuria Lopez-Bigas）	実験医学	2014	国内
Druggable oncogene fusions in invasive mucinous lung adenocarcinoma.	Nakaoku T, Tsuta K, Ichikawa H, Shiraishi K, Sakamoto H, Enari M, Furuta K, Shimada Y, Ogiwara H, Watanabe SI, Nokihara H, Yasuda K, Hiramoto M, Nammo T, Ishigame T, Schetter AJ, Okayama H, Harris CC, Kim YH, Mishima M, Yokota J, Yoshida T, Kohno T*	Clin Cance Res.	2014	国外
A mouse model of KIF5B-RET fusion-dependent lung tumorigenesis.	Saito M, Ishigame T, Tsuta K, Kumamoto K, Imai T, Kohno T	Carcinogenesis	2014	国外
A Three-microRNA signature predicts responses to platinum-Based doublet chemotherapy in patients with lung adenocarcinoma.	Saito M, Shiraishi K, Matsumoto K, Schetter AJ, Ogata-Kawata H, Tsuchiya N, Kunitoh H, Nokihara H, Watanabe SI, Tsuta K, Kumamoto K, Takenoshita S, Yokota J, Harris CC, Kohno T	Clin Cancer Res.	2014	国外

nature genetics

ARTICLES

Trans-ancestry mutational landscape of hepatocellular carcinoma genomes

Yasushi Totoki[1,14], Kenji Tatsuno[2,14], Kyle R Covington[3,14], Hiroki Ueda[2], Chad J Creighton[3,4], Mamoru Kato[1], Shingo Tsuji[2], Lawrence A Donehower[5], Betty L Slagle[5], Hiromi Nakamura[1], Shogo Yamamoto[2], Eve Shinbrot[3], Natsuko Hama[1], Megan Lehmkuhl[3], Fumie Hosoda[1], Yasuhito Arai[1], Kim Walker[3], Mahmoud Dahdouli[3], Kengo Gotoh[2], Genta Nagae[2], Marie-Claude Gingras[3], Donna M Muzny[3], Hidenori Ojima[6], Kazuaki Shimada[7], Yutaka Midorikawa[8], John A Goss[9], Ronald Cotton[9], Akimasa Hayashi[2,10], Junji Shibahara[10], Shumpei Ishikawa[10], Jacfranz Guiteau[9], Mariko Tanaka[10], Tomoko Urushidate[1], Shoko Ohashi[1], Naoko Okada[1], Harsha Doddapaneni[3], Min Wang[3], Yiming Zhu[3], Huyen Dinh[3], Takuji Okusaka[11], Norihiro Kokudo[12], Tomoo Kosuge[7], Tadatoshi Takayama[8], Masashi Fukayama[10], Richard A Gibbs[3], David A Wheeler[3], Hiroyuki Aburatani[2] & Tatsuhiro Shibata[1,13]

Diverse epidemiological factors are associated with hepatocellular carcinoma (HCC) prevalence in different populations. However, the global landscape of the genetic changes in HCC genomes underpinning different epidemiological and ancestral backgrounds still remains uncharted. Here a collection of data from 503 liver cancer genomes from different populations uncovered 30 candidate driver genes and 11 core pathway modules. Furthermore, a collaboration of two large-scale cancer genome projects comparatively analyzed the trans-ancestry substitution signatures in 608 liver cancer cases and identified unique mutational signatures that predominantly contribute to Asian cases. This work elucidates previously unexplored ancestry-associated mutational processes in HCC development. A combination of hotspot *TERT* promoter mutation, *TERT* focal amplification and viral genome integration occurs in more than 68% of cases, implicating *TERT* as a central and ancestry-independent node of hepatocarcinogenesis. Newly identified alterations in genes encoding metabolic enzymes, chromatin remodelers and a high proportion of mTOR pathway activations offer potential therapeutic and diagnostic opportunities.

HCC is the third leading cause of cancer deaths worldwide[1,2]. Epidemiologically, the incidence of HCC shows marked variance across geographical regions and ancestry groups and between the sexes[3]. HCC incidence predominates in East Asia and Africa, and rapid increases in prevalence have occurred in Western countries[2]. Multiple etiological cofactors are associated with liver cancer, and their contributions might additionally differ according to ancestry. Hepatitis B virus (HBV) infection is dominant in East Asia and Africa, whereas hepatitis C virus (HCV) infection among HCC cases is frequent in Japan. Aflatoxin B1 exposure is a strong risk factor of HCC in China and Africa, whereas alcohol intake is a major etiological factor for HCC in Western countries[3–5]. The average male/female ratio for HCC incidence is greater than two, which could be owing to different environmental exposures or hormone levels[6]. Overlapping but partially distinctive epidemiological backgrounds, such as liver fluke infection, were associated with intrahepatic cholangiocarcinoma (IHCC), another type of liver cancer[5]. Here we conducted the first trans-ancestry HCC genome sequencing research under the umbrella of the International Cancer Genome Consortium (ICGC)[7] and The Cancer Genome Atlas (TCGA)[8]. Thus far, this study represents the largest genomic profiling of liver cancers (608 cases) and compares ancestry groups (Japanese, Asian and European) with distinctive etiological cofactors. This genome data set also uncovers an extensive landscape of driver genetic alterations in HCC.

RESULTS

Whole-exome and oncovirome sequencing of liver cancers

As an ICGC liver cancer project, we collected 503 pairs (413 cases in the Japanese cohort and 90 cases in the US cohort) of liver cancers (488 HCC and 15 IHCC) and matched non-cancerous liver tissues

[1]Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan. [2]Genome Science Division, Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan. [3]Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA. [4]Department of Medicine, Baylor College of Medicine, Houston, Texas, USA. [5]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, USA. [6]Division of Molecular Pathology, National Cancer Center Research Institute, Tokyo, Japan. [7]Hepatobiliary and Pancreatic Surgery Division, National Cancer Center Hospital, Tokyo, Japan. [8]Department of Digestive Surgery, Nihon University School of Medicine, Tokyo, Japan. [9]Department of Surgery, Baylor College of Medicine, Houston, Texas, USA. [10]Department of Pathology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [11]Hepatobiliary and Pancreatic Oncology Division, National Cancer Center Hospital, Tokyo, Japan. [12]Hepato-Biliary-Pancreatic Surgery Division, Department of Surgery, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [13]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [14]These authors contributed equally to this work. Correspondence should be addressed to D.A.W. (wheeler@bcm.edu), H.A. (haburata-tky@umin.ac.jp) or T.S. (tashibat@ncc.go.jp).

Received 31 December 2013; accepted 3 October 2014; published online 2 November 2014; doi:10.1038/ng.3126

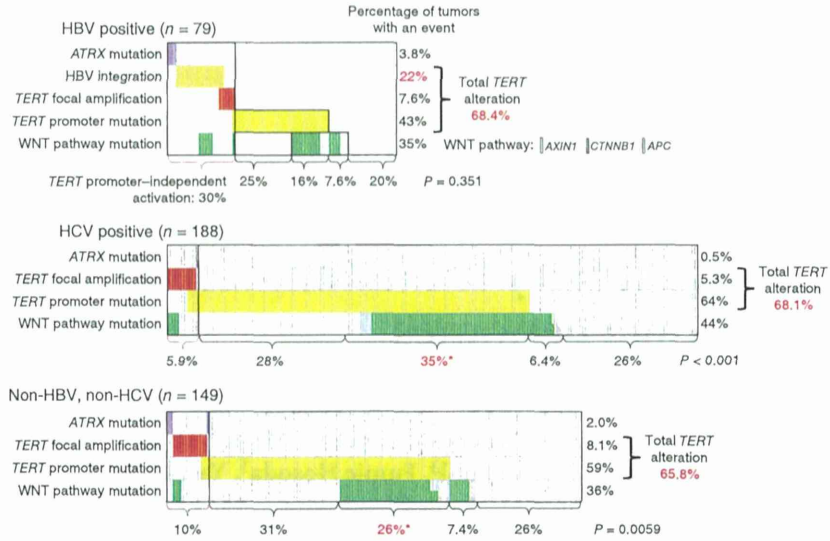

Figure 1 Multiple types of *TERT* alterations in HCC. Mutual exclusivity of HBV genome integration at the *TERT* locus, *TERT* focal amplification and *TERT* promoter mutation in HBV-positive (top), HCV-positive (middle) and non-HBV, non-HCV (bottom) cases. *AXIN1*, *CTNNB1* and *APC* mutations were included as WNT pathway mutations. *TERT* promoter mutation significantly co-occurred with WNT pathway mutation in HBV-negative cases (*$P < 0.001$, χ^2 test). HBV-positive cases without virus capture analysis (41 samples) were excluded (**Supplementary Table 28**).

or blood. This cohort contained 212 HCV-positive, 117 HBV-positive and 150 non-virus cases. The US cohort contained European-ancestry (55%), Asian (defined as US-Asian hereafter; 16%) and African-American (12%) cases. The clinical backgrounds for this cohort are shown in **Supplementary Table 1**.

The exons and surrounding noncoding genomic regions of protein-coding genes were captured in 452 pairs of tumor and non-cancerous liver tissues. Oncoviral genomes, including for HBV, human papillomavirus (HPV-16 and HPV-18) and human T-lymphotrophic virus 1 (HTLV1) (91 kb in total; **Supplementary Table 2**), were also captured in 198 cases. Whole-genome sequencing was conducted in 22 HCC pairs, including 9 exome-sequenced cases, and targeted resequencing of liver cancer genes was carried out for 38 cases. To minimize multicenter study bias due to differences in exome sequencing platform or data analysis pipeline, we optimized the somatic mutation detection algorithms and filtering conditions for three centers using Japanese cohort samples. High concordance (>87%) with a validation rate of >97% in somatic mutation detection was achieved, and substitution patterns among the three centers were consistent (**Supplementary Figs. 1** and **2**). We also confirmed that similar mutation spectra were observed in the same cases in whole-genome sequence and whole-exome sequence (**Supplementary Fig. 3**).

The average mutation rate was 2.8 mutations per megabase, and T>C and C>T substitutions were dominant in this cohort (**Supplementary Fig. 4**). Eight (1.7%) outlier tumors harboring more than 4.3 mutations per megabase showed substitution patterns distinctive from those of other cases and had somatic nonsense or missense mutations in mismatch repair (*MSH3*, *MSH4*, *MSH5* and *MSH6*), DNA polymerase (*POLA1*, *POLK*, *POLE* and *POLL*) or nucleotide excision repair (*ERCC1* and *ERCC2*) genes (**Supplementary Fig. 5**).

Panoramic view of ploidy, copy number and virus integration

We evaluated copy number alteration (CNA) by comparing the sequence depth for paired samples and allelic imbalance in the captured area (**Supplementary Fig. 6**). This digital assessment of CNA and allelic imbalance was consistent with SNP array data in cases analyzed by both methods (**Supplementary Fig. 7**). We also imputed deviation in the allele frequency of heterozygous single-nucleotide variation to predict the tumor purity and ploidy for each sample (H.U., S.Y., K.T. and H.A., unpublished data). A large fraction of cases (28.9%) represented whole-genome duplication with gross chromosomal loss (average ploidy was 3.87, and the average number of CNAs was 11.58) (**Supplementary Fig. 8**), whereas the remainder showed more stable copy number status (average ploidy was 2.08, and the average number of CNAs was 7.56). Tetraploidy was more frequently observed in higher-grade tumors ($P = 0.039$, Fisher's exact test; **Supplementary Fig. 9**).

We observed recurrent arm-level gains (1q, 5p, 6p and 8q) and losses (1p, 4q, 6q, 8p and 17p), as previously described for HCC[9] (**Supplementary Fig. 10**). Recurrent focal amplifications were detected in 25% of cases, including for *TERT* and *CCND1-FGF19*. Homozygous deletions were less frequent events (detected in 17.4% of cases). Recurrent homozygous deletion was observed for 28 genes, including *CDKN2A-CDKN2B*, *MAP2K3* and *PTEN* (**Supplementary Figs. 11** and **12**).

Using paired-end reads mapped to the HBV viral and human genomes, respectively, we detected 628 HBV virus integrations in 68 HBV-positive cases from which viral genomes were captured (9.2 integrations per case) (**Supplementary Table 3**), reflecting a detection rate that was 2–4 times more sensitive than in previous whole-genome sequencing studies[10,11]. Genes close to (less than 10 kb away from) the recurrent HBV integrations included *TERT* ($n = 17$ cases), *KMT2B* (*MLL4*; $n = 6$ cases), and *ALOX5*, *ZFPM2*, *SENP5*, *MYO19* and *RGS22* ($n = 2$ cases each). Recurrent non-genic HBV integrations were observed near the centromere, especially on chromosomes 1p, 8p and 10q. A significant fraction of HBV integrations were colocalized with (less than 500 kb away from) DNA copy number breakpoints (10.7%; $P < 1 \times 10^{-5}$, randomization test) (**Supplementary Figs. 13** and **14**). Despite intimate association between HBV genome integration and CNA breakpoints, the frequency of CNA was not different among the viral subtypes ($P = 0.29$, ANOVA test; **Supplementary Fig. 15** and **Supplementary Table 4**).

Multiple types of *TERT* genetic alteration in HCC

Somatic mutations in the transcriptional regulatory region of the *TERT* gene have been reported in a range of cancers, including HCC[12,13]. By combining captured noncoding sequence data with capillary sequencing validation, we detected *TERT* promoter mutations in 254 cases of the 469 cases analyzed (54% in total). The frequency of these mutations was highest in HCV-positive cases (121/188; 64%), with lower frequencies in non-viral cases (88/149; 59%) and HBV-positive cases (44/120; 37%) (**Supplementary Table 5**). As reported[13], the mutation located 124 bp upstream of the ATG start site (c.–124C>T, on the opposite strand; 93%) was more frequent than the c.–146C>T (4.3%) and c.–57A>C (1.6%) mutations (**Supplementary Table 6**).

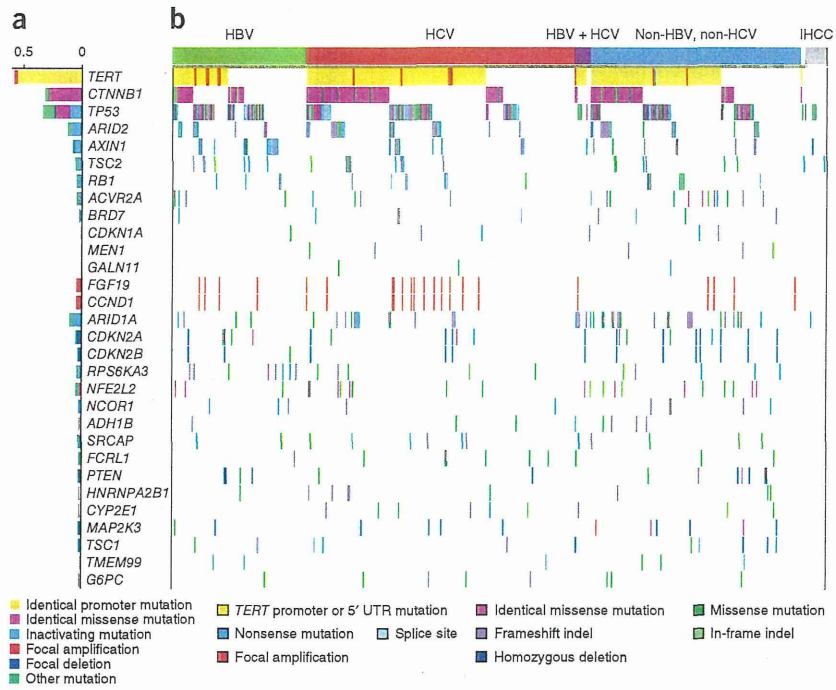

Figure 2 Significant cancer driver genes in HCC. An overview of significant driver genes in HCC. Shown are genes with statistically significant mutations or focal CNAs (**a**) and their alterations in each sample classified by the status of hepatitis virus infection (**b**). Genes were sorted by significant *q* value (**Supplementary Note**).

Additionally, *TERT* focal amplification was detected in 6.7% of the cases in total, and integration of the HBV genome in the *TERT* locus was observed in 22% of HBV-positive samples for which integration was analyzed. *TERT* promoter mutations were mutually exclusive with HBV genome integration in the *TERT* locus in integration-analyzed HBV-positive samples and were almost mutually exclusive with *TERT* focal amplifications, both of which were considered to cause higher *TERT* expression[14] (**Fig. 1**). Alterations of *ATRX* have also been reported to induce telomerase-independent telomere maintenance[15], Altogether, more than 68% of the HCC cases had alterations in either *TERT* or *ATRX*, representing the most frequent molecular event reported (**Supplementary Table 5**). In contrast, no *TERT* promoter mutations were detected in 13 IHCC cases (**Fig. 2**). *TERT* promoter mutations significantly co-occurred with WNT pathway gene alterations, such as *CTNNB1*, *AXIN1* or *APC*, in HCV-positive and non-virus cases, suggesting a cooperative oncogenic activity between *TERT* promoter mutation and the WNT pathway[16] in these subgroups (**Fig. 1**).

Significantly altered genes in HCC

To identify significantly altered genes in HCC, we used a combination of MutSigCV[17], an aggregated somatic alteration method that aggregates somatic substitutions, short indels, homozygous deletions and focal amplifications, and an inactivation bias method that calculates inactivating mutation bias (**Supplementary Fig. 16**, **Supplementary Tables 7–10** and **Supplementary Note**). Furthermore, we eliminated mutated genes that exhibited sequencing center bias and subclone bias as sources of possible false discovery (**Supplementary Tables 11** and **12**). These steps led to a final list of 30 candidate driver genes (**Fig. 2**, **Supplementary Fig. 17** and **Supplementary Tables 13–15**), including 13 that were not recurrently mutated in previous cohorts[18–20] (**Supplementary Table 16**). These 13 genes included *BRD7*, a component of the SWI/SNF nucleosome-remodeling machinery, and *MEN1*, a putative tumor suppressor somatically mutated in neuroendocrine tumors—neither of which has been reported in HCC. Mutations in *TSC2*, *SRCAP* and *NCOR1* have been reported as singletons in other

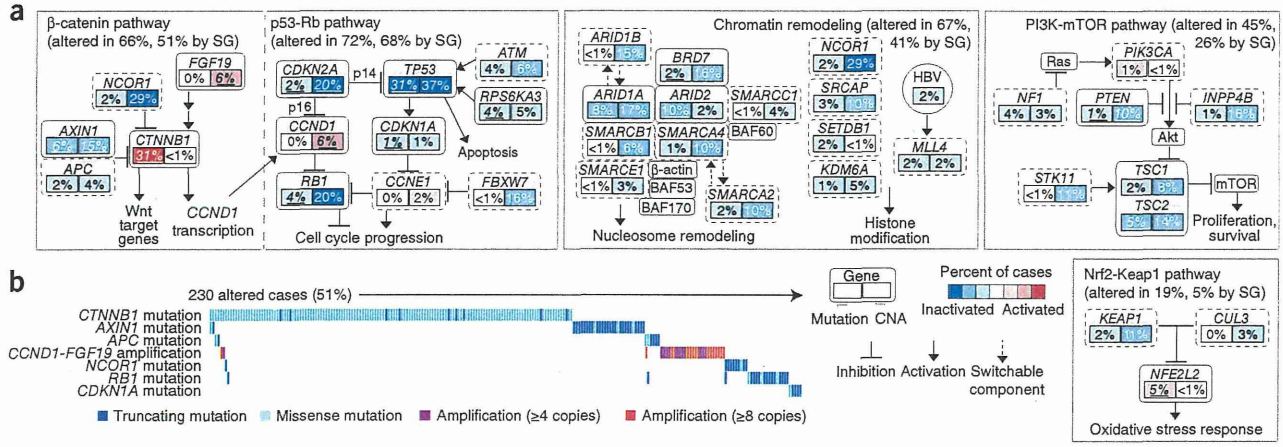

Figure 3 Oncogenic network in HCC. (**a**) Major signaling pathways involving genetic alterations in HCC. Key genes in each pathway are indicated by rectangles, with the percentages of somatic mutations and CNAs shown in the left and right portions of each rectangle, respectively. Significantly altered genes (SG; MutSigCV, $P < 0.05$ or GISTIC, q value < 0.1; percentages are underlined for alterations meeting either criterion) are bounded by solid lines, whereas other key genes in each pathway are bounded by dashed lines. (**b**) Mutual exclusivity plot of genes relevant to the WNT signaling pathway. The plot indicates that somatic mutations in WNT-related genes might contribute to the activation of WNT signaling in over half of all HCCs.

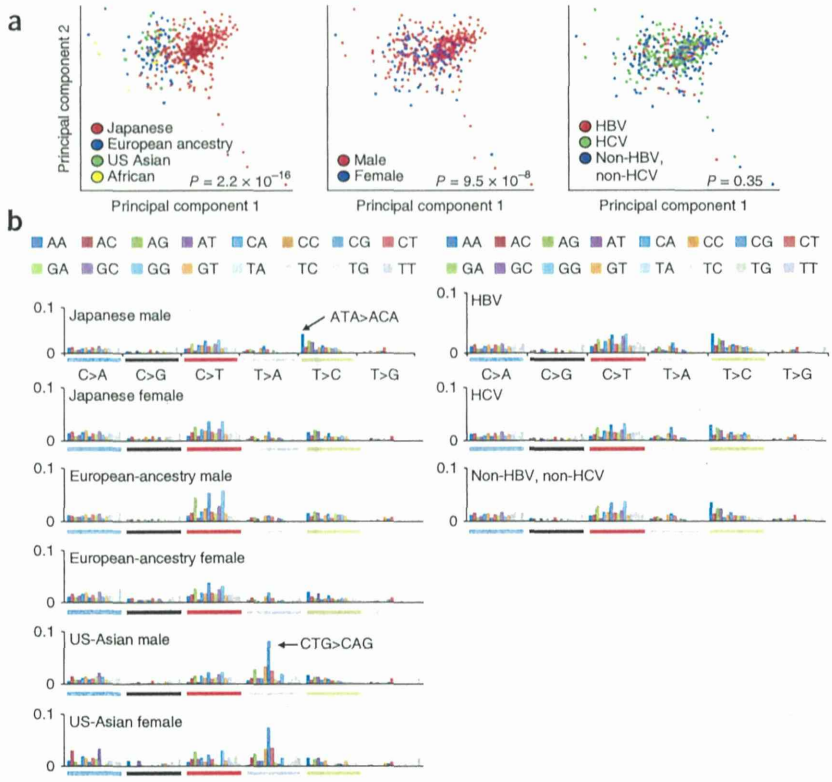

Figure 4 Somatic substitution patterns were associated with ancestry. (**a**) Principal-component analysis of the 96 substitution patterns in the HCC genome by ancestry group (left), sex (middle) and hepatitis virus group (right). (**b**) Average frequency of the 96 substitution patterns in each sample group (ancestry group, sex and virus group). The top legend shows the bases immediately 5′ and 3′ to each substitution. The *y* axis indicates the frequency of the 96 substitution patterns.

studies, but these genes were shown here to be significantly mutated. Some of the difference in results might be attributed to the greatly increased statistical power with our 503-case population, but some of the difference might also reflect contribution from the ancestry composition of the cohorts in this study. Several genes demonstrated differences in mutational frequency among virus subtypes (**Fig. 2b** and **Supplementary Table 17**). *AXIN1* was more frequently mutated in HBV-positive cases in comparison with HCV-positive and non-virus HCC ($P = 0.0055$, Fisher's exact test), indicating that different viral etiologies might activate WNT signaling in distinct ways. *ARID1A* was more frequently altered in non-virus cases ($P = 0.009$).

Alterations of drug target kinases were rarely found in HCC; low-level recurrent mutations of *FGFR2* (mutated in 1.8% of cases), *KIT* (1.3%), *FGFR3* (0.9%), *FGFR1* (0.9%), *JAK1* (0.9%) and *EGFR* (0.4%) and focal amplification of *MET* (0.5%) were detected. The specific mutations in these receptor tyrosine kinases were not generally observed in other cancers, with the exception of two *JAK1* mutations (encoding p.Ser703Ile and p.Leu910Pro substitutions), which were previously observed in a liver cancer sequencing study[20]. The liver has a central role in many metabolic processes. Our study identified recurrent mutations of metabolic enzyme genes in HCC (**Fig. 2b** and **Supplementary Tables 7** and **13**). These included *CYP2E1* (2.0%); *ADH1B* (1.8%), encoding alcohol dehydrogenase 1B; and *G6PC* (1.8%), encoding a glucose-6-phophatase catalytic subunit, whose aberrations could be linked to metabolomic changes in HCC.

Significant oncogenic pathways in HCC

Oncogenic pathways were further explored by aggregating the alterations of each gene within a particular pathway (**Fig. 3a**).

TP53-RB pathway. Inactivation of the tumor-suppressor TP53-RB pathway was a consistent theme in HCC. *TP53* mutations were observed in 31% of tumors, and two genes encoding p53-activating kinases, *ATM* and *RPS6KA3*, were also recurrently mutated. The *RB1* gene was mutated in 4.4% of cases. The *CDKN2A* gene encoding the RB regulator $p16^{INK4A}$ was subject to frequent focal homozygous deletion, and the p53 target and RB regulator *CDKN1A* (encoding $p21^{CIP1}$) was significantly mutated. Overall, 72% of cases had alterations in component genes of one or both of these pathways.

WNT pathway. In addition to activating *CTNNB1* mutations, inactivating mutations were frequently observed in WNT regulators, including *AXIN1* and *APC*. *CCND1* is a key downstream target of WNT signaling[21], and *FGF19* has been shown to activate *CTNNB1* transcriptional functions[22]. Mutual exclusivity of *CTNNB1*, *AXIN1* and *APC* mutations and *CCND1-FGF19* amplification supports the functional role of these genes in altering WNT signaling (**Fig. 3b**). Overall, 66% of HCCs showed WNT pathway–related alterations.

Chromatin and transcription modulators. A large proportion of the genes on the list of significantly mutated genes encoded chromatin modulators or transcriptional regulators. Frequent alterations in *NFE2L2*, encoding a transcriptional regulator that activates antioxidant and cytoprotective target genes[23], and its negative regulators *KEAP1* and *CUL3* (ref. 24) were noted. Also mutated were the nucleosome remodelers *ARID1A*, *ARID2* and *BRD7*, with CNAs and mutations in six additional members of the SWI/SNF complex (**Fig. 3a**), *SRCAP* and the transcriptional corepressor *NCOR1*, both of which have roles in steroid receptor–mediated transcription. These genes displayed primarily inactivating frameshift and nonsense mutations that suggest a tumor-suppressor gene function in HCC (**Supplementary Fig. 18** and **Supplementary Table 9**). *NCOR1* has been shown to directly suppress *CTNNB1* function[25] and exhibits mutual exclusivity for mutations with other WNT pathway genes (**Fig. 3b**). *SRCAP* encodes an Snf2-related CREBBP activator in several pathways, including NOTCH[26] and steroid receptors[27]. Truncating *SRCAP* mutations cause a rare hereditary disease with developmental defects and early-onset tumor formation[28,29], highlighting its potential function as a tumor-suppressor gene.

mTOR-PIK3CA pathway. Recurrent inactivating mutations in *TSC1-TSC2* and activating mutations and copy gain in *PIK3CA* were observed (**Fig. 3a**). Other modulators involved with this pathway, such as *NF1*, *PTEN*, *INPP4B* and *STK11*, were also affected, and, in total, 45% of cases had alterations in the mTOR-PIK3CA pathway. Somatic *TSC1* mutation was reported as a potential predictive biomarker of an mTOR inhibitor[30], and *TSC1*-mutated HCC cell lines showed

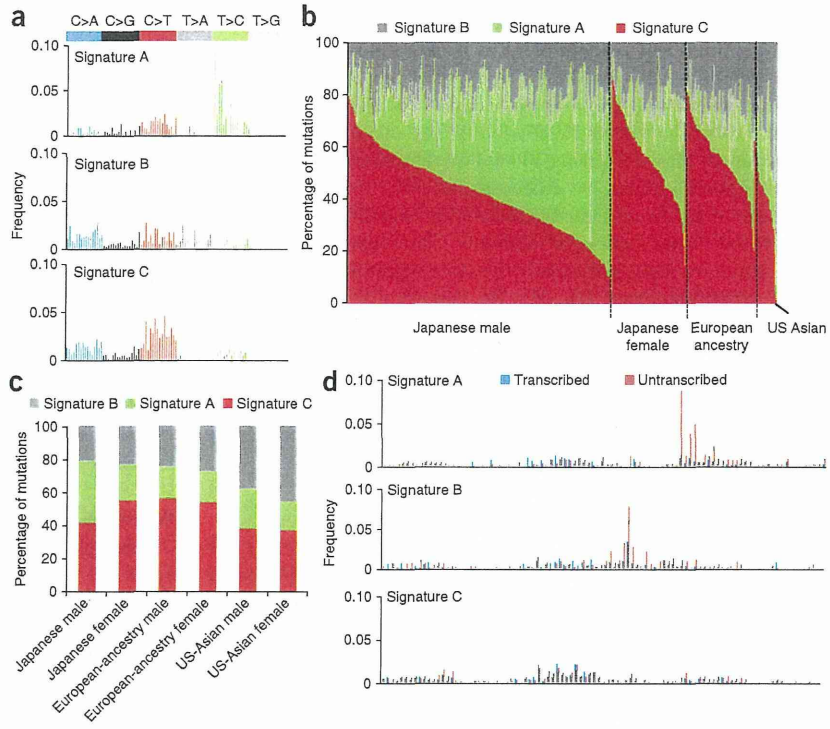

Figure 5 Ancestry-specific mutational signatures with transcriptional strand bias in the HCC genome. (**a**) The 3 mutational signatures in the HCC genome are shown according to the frequencies of 96 substitution types. The *y* axis indicates the frequency of each of the 96 substitution patterns. (**b**) Contribution of the three mutational signatures to each tumor. The *y* axis indicates the percentage of mutations comprised in each signature. The *x* axis indicates tumors classified in each ancestry group and by sex. (**c**) Contribution of the three mutational signatures to tumors from each ancestry group and sex. The *y* axis indicates the percentage of mutations comprised in each signature. (**d**) Transcriptional strand bias in mutational signatures. Each signature is displayed with 192 mutation patterns based on the 96 substitution types with transcriptional strand information. The mutation types are shown on the *x* axis, and the *y* axis indicates the frequency of each of the 192 mutation types contributing to each signature.

higher sensitivity to an mTOR kinase inhibitor (BEZ235) in comparison to cell lines with wild-type *TSC1* (**Supplementary Fig. 19**).

To identify networking among the oncogenic pathways in HCC, we developed a pathway compression algorithm and applied it to the significantly altered genes. We identified 11 core oncogenic network modules in HCC (**Supplementary Table 18**). To visualize these modules in the context of a biological network, we constructed a schematic view of the modules and the additional nodes that can connect them (**Supplementary Fig. 20**). The nodes were typically classified into two types; one type was closely connected to neighboring nodes (with higher value for centrality; **Supplementary Table 19**) and the other type had long-range edges that reached distant nodes, which can be used to measure the effect of each module alteration on the total network. Further comparison of the association between these module alterations and background clinical factors showed that the mTOR module was significantly different ($P < 0.05$, Cochran-Mantel-Haenszel test) in Asian and European-ancestry populations with respect to mutational frequencies (**Supplementary Fig. 21**).

Ancestry-dependent diversity in HCC mutation signatures

Somatic mutation patterns in human cancer are closely associated with epidemiological factors[31–34]; however, their association with ancestry remains unexplored. We integrated genomic data from an additional 105 HCC cases sequenced by TCGA along with the 503 cases sequenced by us (**Supplementary Table 1**) and compared somatic substitution patterns according to epidemiological data and ancestry group. Because mutation patterns in hypermutated cases and IHCC were distinctive (**Supplementary Figs. 4** and **22**), these two groups were excluded from further mutation pattern analysis.

Principal-component analysis of the 96 possible nucleotide triplets, dependent on the bases immediately 5′ and 3′ to each substitution, showed that the constitution of substitution patterns with these triplets was significantly different by ancestry group (Japanese, US Asian and European ancestry; $P = 2.2 \times 10^{-16}$, Wilks' test) and by sex ($P = 9.5 \times 10^{-8}$) (**Fig. 4a**). Notably, substitution patterns were not significantly associated with viral status (HBV, HCV and non-viral, $P = 0.35$; **Fig. 4a** and **Supplementary Fig. 23**). T>C substitutions, particularly in an ATA context, were specifically increased in Japanese male samples, and T>A substitutions (most frequently in a CTG context) were specifically increased in US-Asian male and female samples. The distributions of the frequencies for the 96 substitution types were similar among Japanese female samples and European-ancestry male and female samples (**Fig. 4b**).

We applied non-negative matrix factorization (NMF) analysis to the 96-substitution pattern[33] and identified 3 mutation signatures (HCC signatures A–C; **Fig. 5a** and **Supplementary Fig. 24**). Each signature was composed of context-specific substitutions: HCC signature A was characterized by dominant T>C mutations, especially in an AT(A/G/T) context, whereas HCC signature B contained dominant T>A mutations, with a sharp increase in frequency for a CTG context. HCC signature C contained dominant C>T mutations, especially in an (A/C/G)CG context. The distribution of these signatures was associated with ancestry and sex but not with the virus status (**Supplementary Table 20**). Among the different ancestry groups, HCC signatures A and B more frequently contributed to Japanese male (odds ratio (OR) = 2.2; $P = 0.0025$, Fisher's exact test) and US-Asian (OR = 2.5; $P = 0.00036$) cases, respectively, whereas HCC signature C was common across all ancestry groups and in both sexes (**Fig. 5b,c** and **Supplementary Fig. 25**). Remarkable differences in mutation prevalence between the transcribed and untranscribed strands were observed for T>C substitutions, especially in an AT(A/G/T) context ($P = 7.4 \times 10^{-152}$, χ^2 test), in HCC signature A and for T>A substitutions, especially in a CTG context ($P = 3.3 \times 10^{-8}$), in HCC signature B (**Fig. 5d**). These significant strand biases imply the involvement of transcription-coupled repair, which is tightly associated with known carcinogens in other tumor types[31–34]. There was no significant association between the signature distribution and the *ALDH2* SNP rs671, which is associated with alcohol metabolism and is a more frequent genotype in the Asian population[35] (**Supplementary Table 21**).

To collect large amounts of cancer genome data from different ancestry groups and epidemiological backgrounds, we currently need to combine data from multiple institutes that apply individual analytical platforms. An important caveat in multicenter trans-ancestry analysis has been the possibility that ancestry-specific signatures can be biased by experimental or analytical differences. To avoid this potential bias, we processed the DNA from 99 Japanese HCC cases using the sequencing and analysis pipeline at the United States–based Baylor College of Medicine. Using this data set from a single center, we replicated exactly the same signatures in each population (**Supplementary Fig. 26**). We also examined the distribution of signatures among three centers using Japanese male samples and confirmed that similar distributions were seen among the three centers (**Supplementary Fig. 27**). Furthermore, we analyzed whole-genome sequencing data for 88 Chinese HCC samples[19] and successfully identified HCC signatures B and C in this independent data set (**Supplementary Fig. 28**).

Outcome analysis from mutational signatures

We analyzed the derived NMF signatures to determine whether any signature or signature component was associated with differences in outcome in the HCC cohort. NMF signature values were merged with annotated clinical data. We performed calculations using standardized signature values to control for differences in the mutation rate between the subjects. Multivariate analysis with the Cox proportional hazards model (**Supplementary Fig. 29** and **Supplementary Tables 22–26**) indicated that histological grade, HCC signature B and the interaction with HCC signature A (but not with HCC signature C) were significant predictors of outcome.

DISCUSSION

The present trans-ancestry liver cancer genome study first identified mutational signatures that are independent of hepatitis virus infection and contribute more to the Asian cases than to ones of European ancestry (**Supplementary Tables 27**). One signature, characterized by AT>AC mutations, was predominant in Japanese males, whereas the other, featuring CTG>CAG mutations, was found more frequently in tumors from Asians living in the United States. These correlations may highlight deeper intra-ancestry diversity and/or environmental contributions, and sex bias might further affect downstream target genes and molecular features in HCC[36]. As several genetic loci are associated with individual HCC risk together with HBV and/or HCV infection[37,38], somatic and germline genome interaction might also be important to consider. Notably, these signatures were not evident in IHCC for Japanese cases (data not shown), suggesting that they are unique properties of HCC. The causes of these signatures remain unknown, but skewed transcriptional strand biases in characteristic sequence contexts strongly imply the presence of specific, previously unexplored mutational processes, which profoundly influence tumor genome constitution and behavior.

With 503 cases, this study is the largest liver cancer genome analysis thus far, enabling the formation of a more thorough picture of the mutational landscape of HCC than ever before. In addition to identifying a large number of significantly mutated genes, we have also identified recurrent alterations of 9 of the 14 core genes making up the SWI/SNF complex. We also find a combination of hotspot *TERT* promoter and *ATRX* mutations, along with focal amplification and virus genome integration in the *TERT* locus, in more than 68% of HCC cases regardless of virus subtype. These findings show that *TERT* is a central driver gene and a promising molecular target[39] in HCC. The targeting of high-prevalence mTOR-PIK3CA pathway activation and antiproliferative activity in HCC cells by chemical inhibition should also offer new therapeutic opportunities. In addition, newly identified alterations in the chromatin-remodeling complex and metabolic enzymes are expected to be associated with cancer-specific epigenetic and metabolomic features.

URLs. DNAcopy, http://www.bioconductor.org/packages/2.13/bioc/html/DNAcopy.html; R software, http://www.R-project.org/; R survival package, http://CRAN.R-project.org/package=survival/; HGSC Mercury analysis pipeline, https://www.hgsc.bcm.edu/software/mercury; GRCh38 human reference genome, http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/; BWA2, http://bio-bwa.sourceforge.net/; GATK4, http://www.broadinstitute.org/gatk/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Sequence data have been deposited in the European Genome-phenome Archive (EGA) under accession EGAS00001000389, the ICGC database (http://www.icgc.org/) and the database of Genotypes and Phenotypes (dbGaP) under accession phs000509.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This study was supported by Grants-in-Aid from the Ministry of Health, Labour and Welfare of Japan for the third-term Comprehensive 10-Year Strategy for Cancer Control, grants from the US National Human Genome Research Institute (NHGRI; 5U54HG003273) and National Cancer Institute (NCI; HHSN261201000053C and P30 CA125123), the Program for Promotion of Fundamental Studies in Health Sciences from the National Institute of Biomedical Innovation (NIBIO, Japan) and the National Cancer Center Research and Development Funds (23-A-8, Japan). The National Cancer Center Biobank is supported by the National Cancer Center Research and Development Fund, Japan. The supercomputing resource SHIROKANE was provided by the Human Genome Center at the University of Tokyo (http://sc.hgc.jp/shirokane.html).

AUTHOR CONTRIBUTIONS

Study design: Y.T., K.T., K.R.C., H.U., M.K., D.A.W., H.A. and T.S. Sequencing data generation: K.T., D.M.M., F.H., H. Doddapaneni, H. Dinh, Y.A., K.G., K.W., M.-C.G., T.U., S.O., N.O., M.W. and Y.Z. Data analysis: Y.T., K.T., K.R.C., H.U., M.K., S.T., L.A.D., B.L.S., E.S., S.Y., H.N., M.L., N.H., K.W., K.G., M.D., G.N., D.A.W. and T.S. Statistical analysis: Y.T., K.R.C., H.U., K.T., C.J.C., M.K., S.T. and S.Y. Molecular analysis: Y.A. and T.S. Sample acquisition and clinical data collection: M.-C.G., K.S., Y.M., J.A.G., H.O., A.H., J.S., R.C., J.G., S.I., M.T., T.O., N.K., T.K., T.T. and M.F. Manuscript writing: Y.T., K.T., K.R.C., H.U., C.J.C., L.A.D., B.L.S., M.K., D.A.W., H.A. and T.S. Project oversight: D.A.W., R.A.G., H.A. and T.S.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Jemal, A. *et al.* Global cancer statistics. *CA Cancer J. Clin.* **61**, 69–90 (2011).
2. Forner, A., Llovet, J.M. & Bruix, J. Hepatocellular carcinoma. *Lancet* **379**, 1245–1255 (2012).
3. El-Serag, H.B. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* **142**, 1264–1273 (2012).
4. Yu, J., Shen, J., Sun, T.T., Zhang, X. & Wong, N. Obesity, insulin resistance, NASH and hepatocellular carcinoma. *Semin. Cancer Biol.* **23**, 483–491 (2013).
5. Augustine, M.M. & Fong, Y. Epidemiology and risk factors of biliary tract and primary liver tumors. *Surg. Oncol. Clin. N. Am.* **23**, 171–188 (2014).
6. Tanaka, K., Sakai, H., Hashizume, M. & Hirohata, T. Serum testosterone:estradiol ratio and the development of hepatocellular carcinoma among male cirrhotic patients. *Cancer Res.* **60**, 5106–5110 (2000).
7. International Cancer Genome Consortium. International network of cancer genome projects. *Nature* **464**, 993–998 (2010).

8. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
9. Wang, K. *et al.* Genomic landscape of copy number aberrations enables the identification of oncogenic drivers in hepatocellular carcinoma. *Hepatology* **58**, 706–717 (2013).
10. Sung, W.K. *et al.* Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat. Genet.* **44**, 765–769 (2012).
11. Fujimoto, A. *et al.* Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat. Genet.* **44**, 760–764 (2012).
12. Killela, P.J. *et al. TERT* promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc. Natl. Acad. Sci. USA* **110**, 6021–6026 (2013).
13. Nault, J.C. *et al.* High frequency of telomerase reverse-transcriptase promoter somatic mutations in hepatocellular carcinoma and preneoplastic lesions. *Nat. Commun.* **4**, 2218 (2013).
14. Li, Y. & Tergaonkar, V. Noncanonical functions of telomerase: implications in telomerase-targeted cancer therapies. *Cancer Res.* **74**, 1639–1644 (2014).
15. Heaphy, C.M. *et al.* Altered telomeres in tumors with *ATRX* and *DAXX* mutations. *Science* **333**, 425 (2011).
16. Hoffmeyer, K. *et al.* Wnt/β-catenin signaling regulates telomerase in stem cells and cancer cells. *Science* **336**, 1549–1554 (2012).
17. Lawrence, M.S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
18. Li, M. *et al.* Inactivating mutations of the chromatin remodeling gene *ARID2* in hepatocellular carcinoma. *Nat. Genet.* **43**, 828–829 (2011).
19. Guichard, C. *et al.* Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat. Genet.* **44**, 694–698 (2012).
20. Kan, Z. *et al.* Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res.* **23**, 1422–1433 (2013).
21. Tetsu, O. & McCormick, F. β-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398**, 422–426 (1999).
22. Pai, R. *et al.* Inhibition of fibroblast growth factor 19 reduces tumor growth by modulating β-catenin signaling. *Cancer Res.* **68**, 5086–5095 (2008).
23. Motohashi, H. & Yamamoto, M. Nrf2-Keap1 defines a physiologically important stress response mechanism. *Trends Mol. Med.* **10**, 549–557 (2004).
24. Zhang, D.D., Lo, S.C., Cross, J.V., Templeton, D.J. & Hannink, M. Keap1 is a redox-regulated substrate adaptor protein for a Cul3-dependent ubiquitin ligase complex. *Mol. Cell. Biol.* **24**, 10941–10953 (2004).
25. Song, L.N. & Gelmann, E.P. Silencing mediator for retinoid and thyroid hormone receptor and nuclear receptor corepressor attenuate transcriptional activation by the β-catenin–TCF4 complex. *J. Biol. Chem.* **283**, 25988–25999 (2008).
26. Eissenberg, J.C., Wong, M. & Chrivia, J.C. Human SRCAP and *Drosophila melanogaster* DOM are homologs that function in the Notch signaling pathway. *Mol. Cell. Biol.* **25**, 6559–6569 (2005).
27. Monroy, M.A. *et al.* SNF2-related CBP activator protein (SRCAP) functions as a coactivator of steroid receptor–mediated transcription through synergistic interactions with CARM-1 and GRIP-1. *Mol. Endocrinol.* **17**, 2519–2528 (2003).
28. Hood, R.L. *et al.* Mutations in *SRCAP*, encoding SNF2-related CREBBP activator protein, cause Floating-Harbor syndrome. *Am. J. Hum. Genet.* **90**, 308–313 (2012).
29. Nelson, R.A. *et al.* Floating-Harbor syndrome and intramedullary spinal cord ganglioglioma: case report and observations from the literature. *Am. J. Med. Genet. A.* **149A**, 2265–2269 (2009).
30. Iyer, G. *et al.* Genome sequencing identifies a basis for everolimus sensitivity. *Science* **338**, 221 (2012).
31. Pleasance, E.D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 184–190 (2010).
32. Pleasance, E.D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
33. Alexandrov, L.B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
34. Poon, S.L. *et al.* Genome-wide mutational signatures of aristolochic acid and its application as a screening tool. *Sci. Transl. Med.* **5**, 197ra101 (2013).
35. Goedde, H.W. *et al.* Population genetic studies on aldehyde dehydrogenase isozyme deficiency and alcohol sensitivity. *Am. J. Hum. Genet.* **35**, 769–772 (1983).
36. Keng, V.W. *et al.* Sex bias occurrence of hepatocellular carcinoma in Poly7 molecular subclass is associated with *EGFR*. *Hepatology* **57**, 120–130 (2013).
37. Zhang, H. *et al.* Genome-wide association study identifies 1p36.22 as a new susceptibility locus for hepatocellular carcinoma in chronic hepatitis B virus carriers. *Nat. Genet.* **42**, 755–758 (2010).
38. Kumar, V. *et al.* Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. *Nat. Genet.* **43**, 455–458 (2011).
39. Harley, C.B. Telomerase and cancer therapeutics. *Nat. Rev. Cancer* **8**, 167–179 (2008).