Leveraging Human Genetics to Guide Cancer Drug Development

abstract

Purpose The high attrition rate of cancer drug development programs is a barrier to realizing the promise of precision oncology. We have examined whether the genetic insights from genome-wide association studies of cancer can guide drug development and repurposing in oncology.

Materials and Methods Across 37 cancers, we identified 955 genetic risk variants from the National Human Genome Research Institute-European Bioinformatics Institute genome-wide association study catalog. We linked these variants to target genes using strategies that were based on linkage disequilibrium, DNA three-dimensional structure, and integration of predicted gene function and expression. With the use of the Informa Pharmaprojects database, we identified genes that are targets of unique drugs and assessed the level of enrichment that would be afforded by incorporation of genetic information in preclinical and phase II studies. For targets not under development, we implemented machine learning approaches to assess druggability.

Results For all preclinical targets incorporating genetic information, a 2.00-fold enrichment of a drug being successfully approved could be achieved (95% CI, 1.14- to 3.48-fold; $P = .02$). For phase II targets, a 2.75-fold enrichment could be achieved (95% CI, 1.42- to 5.35-fold; $P < .001$). Application of genetic information suggests potential repurposing of 15 approved nononcology drugs.

Conclusion The findings illustrate the value of using insights from the genetics of inherited cancer susceptibility discovery projects as part of a data-driven strategy to inform drug discovery. Support for cancer germline genetic information for prospective targets is available online from the Institute of Cancer Research.

Clin Cancer Inform. © 2018 by American Society of Clinical Oncology

INTRODUCTION

The high attrition rate of drug development programs represents a significant barrier to fully realizing the vision of precision oncology.¹ The failure of preclinical model systems to adequately predict efficacy in humans is leading drug developers to seek additional sources of evidence to inform decisions about which targets to pursue.^{2,3}

Since the completion of the Human Genome Project, rapid progress has been made in identifying inherited genetic variants that influence cancer risk through genome-wide association studies (GWASs) and large-scale sequencing projects.⁴ GWASs now have been performed for most common malignancies and many rare tumor types, and more than 900 genetic variants have been robustly demonstrated to influence risk.⁴

The insights from these GWASs potentially offer an additional mechanism for selecting drug

targets and indications, both key requirements in drug discovery. Risk single-nucleotide polymorphisms (SNPs) in or near a gene that may associate with the activity or expression of the encoded proteins can, therefore, be used as a tool to infer the effect of pharmacologic action on the same protein in a trial. Specifically, by extension, disease-associated SNPs identified by GWASs can be interpreted explicitly as a source of randomized human evidence to aid drug target identification and validation.

Several examples illustrate the application of human genetics to inform drug discovery by using knowledge of variation in genes associated with disease risk. These include the targeting of 3-hydroxy-3-methylglutaryl-coenzyme A reductase by statins for the treatment of coronary heart disease⁵ and ustekinumab, a monoclonal inhibitor of interleukin-12 and -23 used to treat inflammatory bowel disease.⁶ With the use of GWAS data for 37 cancers, we have examined the potential for human genetics to guide cancer

Ben Kinnersley
Amit Sud
Elizabeth A. Coker
Joseph E. Tym
Patrizio Di Micco
Bissan Al-Lazikani
Richard S. Houlston

Author affiliations and support information (if applicable) appear at the end of this article.

Corresponding author:

Ben Kinnersley, PhD,
Division of Genetics and
Epidemiology, The Institute
of Cancer Research,
15 Cotswold Rd, Sutton,
London SM2 5NG, UK;
Twitter: @HoulstonLab_
ICR; e-mail: ben.
kinnersley@icr.ac.uk.

drug development and the repurposing of current approved drugs.

MATERIALS AND METHODS

Compilation of GWAS Data

To curate cancer risk SNPs identified by GWAS, we queried the National Human Genome Research Institute-European Bioinformatics Institute GWAS catalog.^{7,8} We imposed a number of quality control metrics, filtering by association $P < 5 \times 10^{-8}$, and included only SNPs associated with the cancer rather than with another cancer-related phenotype, such as progression. We also manually added SNPs from recent cancer GWASs that had not been added yet to the catalog (Data Supplement). We considered GWAS associations irrespective of ethnicity. Gene transcript information, including gene annotations and transcript start sites for human build 37, were obtained from the Ensembl Genes 89 data set.⁹

Linking Risk SNPs to Target Genes

To the extent that they have been deciphered, most GWAS risk SNPs map to noncoding regions of the genome and influence gene regulation. Because spatial proximity between specific genomic regions and chromatin looping interactions are central for the regulation of gene expression, the three-dimensional structure of DNA means that the gene that is closest to the risk SNP does not always necessarily equate to the target gene. However, regulatory effects, and hence target genes, generally are confined within topologically associated domains (TADs) of the genome. To link risk SNPs to target genes, we therefore adopted three strategies.

For linkage disequilibrium (LD)-based annotation, an approach similar to that adopted by Finan et al¹⁰ was undertaken. For each cancer risk SNP, correlated SNPs were obtained for European (CEU [Northern Europeans From Utah]), east Asian (CHB [Han Chinese in Beijing, China]), and African (YRI [Yoruba in Ibadan, Nigeria]) populations from 1000 Genomes Project phase 3 using the LDlink Web application.^{11,12} LD boundaries were designated by the smallest and largest genomic location of SNPs correlated ($r^2 = 0.1$ to 0.9) with the reported cancer risk SNP. For SNPs where LD information could not be obtained, the boundaries were taken as 2.5 kb on either side of the SNP genomic position.

Gene transcription start sites were then mapped to these LD boundaries.

TAD boundaries that encompass each risk locus were based on H1 human embryonic stem cells that were obtained from Schmitt et al.¹³ These data make use of Hi-C data described in Dixon et al.¹⁴ TAD boundaries were identified using the insulation score approach proposed by Crane et al¹⁵ at 40-kb resolution.

To explore target gene prioritization further, we used Data-driven Expression Prioritized Integration for Complex Traits (DEPICT; <https://data.broadinstitute.org/mpg/depict>),¹⁶ an integrative tool that is based on predicted gene function, prioritizes the most likely target genes of risk SNPs, and uses gene expression data from multiple sources. SNP associations were pruned to a set of independent signals by r^2 greater than 0.05 in YRI, CEU, and CHB populations, also retaining SNPs for which LD metrics could not be obtained. We considered all target genes with a false discovery rate Q less than 0.05 as well as the top gene per SNP.

Finally, as an adjunct to our GWAS-based analysis, we also considered the classic cancer susceptibility genes (CSGs) whose mutation in the germline is responsible for the various Mendelian forms of cancer. These were obtained from the Catalog of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census 14.¹⁷

Genetic Association Enrichment for Approved Drugs

Data on the status of drug-target combinations along the various stages of drug development from preclinical through regulatory approval were obtained by interrogation of the Informa Pharmaprojects database.¹⁸ In addition to drugs assessed by Pharmaprojects, cancer drugs approved for use in CSG carriers also were considered. Drugs with a specific indication for symptom control only were excluded. Records were retained if target genes could be mapped unambiguously to Human Genome Organisation Gene Nomenclature Committee identifiers.¹⁹ We assessed whether drug targets with supporting genetic evidence were more likely to be approved in the drug development pipeline by constructing a two-by-two table of genes and counts that correspond to whether a gene product has genetic support as a drug target at

respective stages of development (eg, a comparison of approved drugs with those that only reach preclinical stages). Test of association was Fisher's exact test, with the Wald test used to quantify effect size and 95% CIs. $P = .05$ (two sided) was considered statistically significant. All statistical calculations were performed using R 3.2 software (<https://cran.r-project.org/bin/windows/base/old/3.2.0>).

Druggability Annotation of Target Genes

Targets of Food and Drug Administration–approved drugs were obtained from Santos et al.²⁰ Genes were filtered for protein-coding genes, and the canSAR v4 Cancer Protein Annotation Tool²¹ was used to identify proteins with greater than 95% sequence homology to existing drug targets. The Cancer Protein Annotation Tool also was used to extract structure- and ligand-based druggability assessments from canSAR.²² Network-based druggability scores for proteins were based on Mitsopoulos et al.²³

Finally, we assessed all 355,305 active compounds identified by canSAR against their targets using Probe Miner,²⁴ which catalogs more than 1.8 million compounds for their suitability as chemical tools against 2,220 UniProt-defined human targets (<http://probeminer.icr.ac.uk/#/download>).

RESULTS

Linking Risk SNPs to Target Genes

Across 37 cancers, we identified 955 risk loci. To link sentinel risk SNPs to respective target genes, we first considered genes within regions of LD to which risk SNPs mapped by imposing a range of r^2 thresholds. Next, we considered all genes that localized within the risk SNP–defined TAD boundaries. Finally, we based linkage on the gene prioritization approach implemented in DEPICT.¹⁶ These three approaches yielded between 394 and 7,379 protein-coding target genes (Fig 1; Data Supplement).

Genetic Association Enrichment for Approved Drugs

By interrogating the Informa Pharmaprojects database, we identified 1,706 unique genes that were the target of 3,435 unique therapeutic agents for cancer (Data Supplement). These

were grouped according to the farthest point reached across five stages of the drug development pipeline: preclinical (ie, in vitro and in vivo dosing and toxicity assessment), phase I (safety and dosage), phase II (efficacy and adverse effects), phase III and preregistration (efficacy and monitoring of adverse reactions), and approved.

We first considered all targets from the preclinical stage and assessed the level of enrichment for being successfully approved conferred by genetic information. All the methods linking SNPs to target genes provided evidence for enrichment. For the LD-based assessment, enrichment was strongly correlated with r^2 values; imposition of an r^2 greater than 0.9 resulted in 2.00-fold improvement in targeting of preclinical drugs (95% CI, 1.14- to 3.48-fold; $P = .02$; Fig 2A; Table 1). The comparative enrichment associated with COSMIC-cataloged CSGs was 6.61-fold (95% CI, 3.17- to 13.78-fold; $P < .001$; Fig 2A; Table 1).

We reasoned that a target's failure to progress along the preclinical and phase I stages often is unrelated to efficacy; therefore, we next considered all targets from phase II and above and assessed the degree of enrichment for approval conferred by genetic information. As with the analysis of preclinical targets, incorporation of genetic association information led to enrichment for approval (Fig 2B; Table 2). The strongest enrichment from the LD-based approach was attained after imposing an r^2 greater than 0.9, which was associated with a significant 2.75-fold difference (95% CI, 1.42- to 5.35-fold; $P < .001$; Fig 2B). The comparative enrichment associated with COSMIC-cataloged CSGs was 5.72-fold (95% CI, 2.35- to 13.89-fold; $P < .001$).

Potential for Repurposing Noncancer Drugs

To explore the application of genetics to inform drug repurposing, we first identified approved drugs used in the treatment of nononcologic disease. We then examined discordant pairings of drug indications and cancer associations. We identified 15 genes for which an approved drug currently is available with genetic support (Table 3). Notable examples are *TGFB1* at 19q13.2, where a targeted drug is used in the treatment of rheumatoid arthritis and is the site of an association

Fig 1. Summary of analytical strategy. DEPICT, Data-driven Expressed Prioritized Integration for Complex Traits; SNP, single-nucleotide polymorphism; TAD, topologically associating domain.

with colorectal cancer risk²⁵; *VDR* at 12q13.11, which is targeted by drugs that treat osteoporosis and is a risk locus for prostate cancer²⁶; *TYR* at 11q14.3, which is the target of an approved drug used in the treatment of skin disorders and is the site of a risk locus for melanoma,²⁷ squamous cell carcinoma,²⁸ and basal cell carcinoma²⁹; and *PTGIR* at 19q13.32, which is targeted by a drug used in the treatment of transplant rejection and peripheral vascular disease and is the site of a chronic lymphocytic leukemia risk locus.³⁰

Availability of Cancer Germline Genetic Information

Supporting cancer germline genetic information for prospective targets is available from canSAR²¹ at <https://cansar.icr.ac.uk> (Fig 3). For each UniProt identifier, a report has been generated that

details whether the given gene has been annotated as containing cancer-causing germline mutations by the COSMIC germline cancer gene census³¹ as well as whether any variants from cancer GWASs map to the gene (Fig 3).

DISCUSSION

These findings support the potential of human genetics to guide the identification of drug targets by addressing a productivity-limiting step in drug development and a bottleneck to realizing the vision of precision oncology. Specifically, we have demonstrated that knowledge of CSGs identified by GWASs can be used to maximize discovery of likely preclinical and phase II targets, thereby empowering drug development programs. Our analysis benefits from the larger number of risk loci for cancer that have been

Fig 2. Enrichment of approved cancer drug targets that incorporates genetic evidence relative to preclinical (A) and phase II (B) targets. Data are based on [Tables 1](#) and [2](#). CSG, cancer susceptibility gene; TAD, topologically associated domain.

identified over recent years, thereby providing greater power than earlier studies.¹

Significant enrichment of preclinical and phase II targets also was shown by incorporating information on the classic CSGs. Given that many of the CSGs are somatically mutated, these targets may have already directly influenced recent drug

development programs. Indeed, we observed a highly significant enrichment for CSGs being selected for preclinical validation per se (odds ratio, 11.37; 95% CI, 7.44 to 17.37; $P < .001$), which is greater than that afforded to genes simply implicated by GWAS (r^2 greater than 0.9) targets (odds ratio, 1.44; 95% CI, 1.14 to 1.81; $P = .003$).

Table 1. Enrichment of Approved Cancer Drug Targets Supported by Genetic Evidence Relative to Preclinical Targets

Method	Preclinical Drug, No.				Enrichment (preclinical v approved)	
	Approved With Genetic Support	Not Approved With Genetic Support	Approved With No Genetic Support	Not Approved With No Genetic Support	P	OR (95% CI)
TAD	49	264	66	457	.22	1.28 (0.86 to 1.92)
$r^2 > 0.1$	36	186	79	535	.21	1.31 (0.85 to 2.01)
$r^2 > 0.2$	28	141	87	580	.26	1.32 (0.83 to 2.11)
$r^2 > 0.3$	26	125	89	596	.19	1.39 (0.86 to 2.25)
$r^2 > 0.4$	24	109	91	612	.13	1.48 (0.90 to 2.42)
$r^2 > 0.5$	22	92	93	629	.08	1.62 (0.97 to 2.70)
$r^2 > 0.6$	21	89	94	632	.10	1.59 (0.94 to 2.67)
$r^2 > 0.7$	19	77	96	644	.08	1.65 (0.96 to 2.86)
$r^2 > 0.8$	19	69	96	652	.03	1.87 (1.08 to 3.25)
$r^2 > 0.9$	19	65	96	656	.02	2.00 (1.14 to 3.48)
COSMIC germline	15	16	100	705	< .001	6.61 (3.17 to 13.78)
DEPICT	8	33	107	688	.25	1.56 (0.70 to 3.46)

NOTE. Enrichment was calculated by constructing a two-by-two table of genes and counts that correspond to whether a gene product has genetic support at the respective stages of drug development (ie, approved v preclinical). Test of association was Fisher's exact test, with the Wald test used to quantify effect size and 95% CIs. $P = .05$ (two sided) was considered statistically significant.

Abbreviations: COSMIC, Catalog of Somatic Mutations in Cancer; DEPICT, Data-driven Expressed Prioritized Integration for Complex Traits; OR, odds ratio; TAD, topologically associated domain.

We used a number of methods to map target genes to cancer risk SNPs by incorporating LD blocks, TAD regions, and gene expression. We found that genes implicated by the LD r^2 greater than 0.9 method showed the greatest enrichment for drug approval. Although compatible with

Table 2. Enrichment of Approved Cancer Drug Targets Supported by Genetic Evidence Relative to Phase II Targets

Method	Phase II Drug, No.				Enrichment (phase II v approved)	
	Approved With Genetic Support	Not Approved With Genetic Support	Approved With No Genetic Support	Not Approved With No Genetic Support	P	OR (95% CI)
TAD	49	100	66	213	.04	1.58 (1.02 to 2.45)
$r^2 > 0.1$	36	72	79	241	.10	1.53 (0.95 to 2.45)
$r^2 > 0.2$	28	50	87	263	.07	1.69 (1.00 to 2.85)
$r^2 > 0.3$	24	46	89	267	.06	1.70 (0.99 to 2.90)
$r^2 > 0.4$	22	40	91	273	.05	1.80 (1.03 to 3.15)
$r^2 > 0.5$	21	32	93	281	.02	2.08 (1.15 to 3.75)
$r^2 > 0.6$	19	30	94	283	.02	2.11 (1.15 to 3.86)
$r^2 > 0.7$	19	26	96	287	.02	2.18 (1.16 to 4.12)
$r^2 > 0.8$	19	23	96	290	< .001	2.50 (1.30 to 4.78)
$r^2 > 0.9$	19	21	96	292	< .001	2.75 (1.42 to 5.35)
COSMIC germline	15	8	100	305	< .001	5.72 (2.35 to 13.89)
DEPICT	8	9	107	304	< .001	2.52 (0.95 to 6.71)

NOTE. Enrichment was calculated by constructing a two-by-two table of genes and counts that correspond to whether a gene product has genetic support at the respective stages of drug development (ie, approved v phase II). Test of association was Fisher's exact test, with the Wald test used to quantify effect size and 95% CIs. $P = .05$ (two sided) was considered statistically significant.

Abbreviations: COSMIC, Catalog of Somatic Mutations in Cancer; DEPICT, Data-driven Expressed Prioritized Integration for Complex Traits; OR, odds ratio; TAD, topologically associated domain.

Table 3. Opportunities for Drug Repurposing Informed by Germline Cancer Genetics

Gene	Entrez	Additional Genes Targeted by Drug	Disease	Locus	Cancer Type
<i>ALOX5</i>	240		Asthma, COPD	10q11.21	Prostate cancer
<i>CFTR</i>	1080		Cystic fibrosis, diarrhea, short-bowel syndrome, irritable bowel syndrome, diarrhea-predominant infection, GI tract infection, HSV infection, HIV/AIDS	7q31.2	Barrett's esophagus, esophageal adenocarcinoma
<i>CLCN2</i>	1181		Chronic constipation, irritable bowel syndrome, GI motility dysfunction	3q27.1	Esophageal adenocarcinoma
<i>CRHR1</i>	1394	<i>CRHR2</i> (1395)	Anxiety, unspecified insomnia	17q21.31	Ovarian cancer in <i>BRCA1</i> carriers
<i>DDC</i>	1644		Parkinson's disease	7p12.1	Childhood ALL
<i>GABBR1</i>	2550		Spasticity, multiple sclerosis, alcohol addiction, cerebral palsy, spinal cord injury, dystonia	6p22.1	Barrett's esophagus, esophageal adenocarcinoma
<i>GBA</i>	2629		Gaucher's disease	1q22	Gastric adenocarcinoma
<i>INSR</i>	3643		Diabetes type 1, diabetes type 2	19p13.2	Renal cell carcinoma, differentiated thyroid cancer
<i>PDE4D</i>	5144	<i>PDE4A</i> (5141) <i>PDE4B</i> (5142) <i>PDE4C</i> (5143)	COPD, asthma, nonalcoholic steatohepatitis, eczema, Alzheimer's disease, schizophrenia, rhinitis, psoriasis	5q12.1	Esophageal cancer, breast cancer
<i>PLG</i>	5340		Venous thrombosis, myocardial infarction, pulmonary thrombosis	6q26	Prostate cancer
<i>PTGIR</i>	5739		Pulmonary hypertension, transplant rejection, peripheral vascular disease, limb ischemia	19q13.32	CLL
<i>SLC6A3</i>	6531		Depression, CNS diagnosis, ADHD	5p15.33	Pancreatic cancer
<i>TGFB1</i>	7040		Wound healing, conjunctivitis, asthma, eczema, rhinitis, rheumatoid arthritis, hyperuricemia, multiple sclerosis, restenosis	19q13.2	Colorectal cancer
<i>TYR</i>	7299		Skin disorder	11q14.3	Melanoma, squamous cell carcinoma, basal cell carcinoma
<i>VDR</i>	7421		Osteoporosis, keratosis, secondary hyperparathyroidism, psoriasis, osteodystrophy, hypophosphatemia, palmoplantar pustulosis, ichthyosis	12q13.11	Prostate cancer

NOTE. Targets annotated to cancer risk single-nucleotide polymorphisms by $r^2 > 0.8$ were assessed for overlap with approved nononcology drugs.

Abbreviations: ADHD, attention deficit hyperactivity disorder; ALL, acute lymphoblastic leukemia; CLL, chronic lymphocytic leukemia; COPD, chronic obstructive pulmonary disease; HSV, herpes simplex virus.

the functional basis of many GWAS associations resulting from the most proximal genes, this does not preclude the possibility of longer-range tissue-specific mechanisms that are less amenable to detection by our approach. Therefore, future endeavors of this kind likely will benefit from more-detailed experimental investigation of the biologic mechanism that underpins cancer risk loci. Although the TAD-based strategy is likely to be always beset by the issue of capturing too many genes, strategies that are based on integration of GWASs and multiomics per DEPICT¹⁶ are likely to

improve, which will make them attractive sources of genetic information. To investigate regulatory interactions across all cancer risk loci, we made use of publicly available Hi-C data from human embryonic stem cells, and took note of the observation of Dixon et al³³ that TAD boundaries are relatively stable across cell types. However, the increasing availability of tissue- and cancer-specific Hi-C data is likely to improve efforts to identify target genes of specific cancer risk regions.

In concert with our primary analysis, we identified a number of possible opportunities for drug

Fig 3. Integration of cancer germline genetics information into canSAR. Available germline genetic evidence can be searched for by target through the Institute of Cancer Research,^{22,32} where P23458 is the UniProt identifier (ID) for the target of interest. 3D, three dimensional; DEPICT, Data-driven Expressed Prioritized Integration for Complex Traits; FAQ, frequently asked question; HGNC, Human Genome Organisation Gene Nomenclature Committee; LD, linkage disequilibrium; PMID, PubMed identifier; RAF, risk allele frequency; RNAi, RNA interference; SNP, single-nucleotide polymorphism; TAD, topologically associated domain.

repurposing as informed by cancer germline genetics. These extend the potential of preexisting therapies and highlight that pathways subverted by cancers also may be altered in other diseases.

For pragmatic purposes, we considered all cancers, with the assumption that generic effects exist at least across some cancers, to maximize study power. We acknowledge, however, that this is, in essence, crude because certain cancer subtypes can show specific associations with risk SNPs, which reflects differences in their biology, such as estrogen receptor-positive and -negative breast cancers^{34,35} and combinations of 1p/19q codeletion, TERT promoter, and *IDH* mutation in glioma.^{36,37} The future availability of larger data sets that will afford the identification of additional risk SNPs will open up the possibility of fine-tuned analyses. In addition, we assume that cancer risk variants act directly to influence cancer initiation or progression, but this does not preclude the existence of a limited subset that may have indirect mechanisms, such as at 15q25.1, where the association with lung cancer is likely to be due to smoking.³⁸

One caveat to using all forms of germline genetics as a mechanism for prioritization of drug development is the assumption that susceptibility per se is also reflective of progression, which may not always be the case. As with other studies, we have used drug approval as a surrogate for drug efficacy. This assumption, however, will only make our estimates more conservative. We also acknowledge our lack of inclusion of generic drugs, but because the majority of these have a broad range of targets, we do not regard this as a significant impact on our findings. While considering the extent to which cancer genes implicated by GWASs that are not currently in the drug development pipeline might represent good candidates, we performed multifaceted druggability analyses that incorporated assessments of the three-dimensional structures of the target protein and any associated protein complexes, chemical properties of known ligands of the target, and the target's position and role within the human interactome and ranked target-indication pairings by criteria, including novelty relative to existing targets and predicted attrition risk (Data Supplement). Of 1,292 genes annotated to GWAS SNPs by r^2 greater than 0.9,

977, 486, and 1,287 had druggability assessments by network, structure, and ligand-based prediction, respectively. Of note, 29 of these can be targeted by existing high-quality probes and thus, represent good candidates for being prioritized in future studies.

In conclusion, we demonstrate that enrichment for targets implicated by cancer risk variants are more successful in the drug development pipeline, which provides a rationale for germline genetics that empower cancer drug discovery. The mapping of approved drug targets back to cancer GWAS signals enables the identification

of both novel drug targets and patient populations. To benefit the wider community, the cancer germline information used in this study is available from the Institute of Cancer Research at <https://cansar.icr.ac.uk>.²² Collectively, our findings show the value of incorporating information from germline cancer genetics as part of interdisciplinary, data-driven approaches to inform drug discovery in oncology.

DOI: <https://doi.org/10.1200/CCI.18.00077>

Published online on ascopubs.org/journal/cci on November 21, 2018.

AUTHOR CONTRIBUTIONS

Conception and design: Ben Kinnersley, Amit Sud, Elizabeth A. Coker, Bissan Al-Lazikani, Patrizio Di Micco, Richard S. Houlston

Administrative support: Richard S. Houlston

Collection and assembly of data: Ben Kinnersley, Amit Sud, Elizabeth A. Coker, Patrizio Di Micco

Data analysis and interpretation: Ben Kinnersley, Amit Sud, Elizabeth A. Coker, Joseph E. Tym, Patrizio Di Micco, Richard S. Houlston

Manuscript writing: All authors

Final approval of manuscript: All authors

Accountable for all aspects of the work: All authors

matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/ifc.

Ben Kinnersley

No relationship to disclose

Amit Sud

No relationship to disclose

Elizabeth A. Coker

No relationship to disclose

Joseph E. Tym

No relationship to disclose

Patrizio Di Micco

No relationship to disclose

Bissan Al-Lazikani

No relationship to disclose

Richard S. Houlston

No relationship to disclose

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject

Affiliations

All authors: The Institute of Cancer Research, London, United Kingdom.

Support

Supported by a Team Science project grant from the Institute of Cancer Research and by Cancer Research UK. A.S. is supported by charitable funds from the Royal Marsden Hospital.

Prior Presentation

Presented at the American Association for Cancer Research Annual Meeting 2018, Chicago, IL, April 14-18, 2018.

REFERENCES

1. Nelson MR, Tipney H, Painter JL, et al: The support of human genetic evidence for approved drug indications. *Nat Genet* 47:856-860, 2015
2. Shih HP, Zhang X, Aronov AM: Drug discovery effectiveness from the standpoint of therapeutic mechanisms and indications. *Nat Rev Drug Discov* 17:78, 2018
3. Workman P, Draetta GF, Schellens JHM, et al: How much longer will we put up with \$100,000 cancer drugs? *Cell* 168:579-583, 2017

4. Sud A, Kinnersley B, Houlston RS: Genome-wide association studies of cancer: Current insights and future perspectives. *Nat Rev Cancer* 17:692-704, 2017
5. Taylor F, Huffman MD, Macedo AF, et al: Statins for the primary prevention of cardiovascular disease. *Cochrane Database Syst Rev* (1):CD004816, 2013
6. Feagan BG, Sandborn WJ, Gasink C, et al: Ustekinumab as induction and maintenance therapy for Crohn's disease. *N Engl J Med* 375:1946-1960, 2016
7. Welter D, MacArthur J, Morales J, et al: The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 42:D1001-D1006, 2014
8. European Molecular Biology Laboratory-Bioinformatics Institute: GWAS catalog, 2017. <http://www.ebi.ac.uk/gwas>
9. European Molecular Biology Laboratory-Bioinformatics Institute: GRCh37, 2017. <http://grch37.ensembl.org/biomart/martview>
10. Finan C, Gaulton A, Kruger FA, et al: The druggable genome and support for target identification and validation in drug development. *Sci Transl Med* 9:9, 2017
11. Machiela MJ, Chanock SJ: LDlink: A Web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31:3555-3557, 2015
12. National Cancer Institute, Division of Cancer Epidemiology & Genetics: Welcome to LDlink. <https://analysistools.nci.nih.gov/LDlink>
13. Schmitt AD, Hu M, Jung I, et al: A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Reports* 17:2042-2059, 2016
14. Dixon JR, Jung I, Selvaraj S, et al: Chromatin architecture reorganization during stem cell differentiation. *Nature* 518:331-336, 2015
15. Crane E, Bian Q, McCord RP, et al: Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* 523:240-244, 2015
16. Pers TH, Karjalainen JM, Chan Y, et al: Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* 6:5890, 2015
17. Wellcome Sanger Institute: Cancer Gene Census. <https://cancer.sanger.ac.uk/census>
18. Informa: Pharma intelligence. <https://pharmaintelligence.informa.com>
19. HUGO Gene Nomenclature Committee: Home page. <https://www.genenames.org>
20. Santos R, Ursu O, Gaulton A, et al: A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* 16:19-34, 2017
21. Tym JE, Mitsopoulos C, Coker EA, et al: canSAR: An updated cancer research and drug discovery knowledgebase. *Nucleic Acids Res* 44:D938-D943, 2016
22. Institute of Cancer Research: canSAR Version 4.0, 2018. <https://cansar.icr.ac.uk>
23. Mitsopoulos C, Schierz AC, Workman P, et al: Distinctive behaviors of druggable proteins in cellular networks. *PLOS Comput Biol* 11:e1004597, 2015
24. Antolin AA, Tym JE, Komianou A, et al: Objective, quantitative, data-driven assessment of chemical probes. *Cell Chem Biol* 25:194-205, 2017
25. Zhang B, Jia WH, Matsuda K, et al: Large-scale genetic study in East Asians identifies six new loci associated with colorectal cancer risk. *Nat Genet* 46:533-542, 2014
26. Al Olama AA, Kote-Jarai Z, Berndt SI, et al: A meta-analysis of 87,040 individuals identifies 23 new susceptibility loci for prostate cancer. *Nat Genet* 46:1103-1109, 2014
27. Bishop DT, Demenais F, Iles MM, et al: Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet* 41:920-925, 2009
28. Asgari MM, Wang W, Ioannidis NM, et al: Identification of susceptibility loci for cutaneous squamous cell carcinoma. *J Invest Dermatol* 136:930-937, 2016

29. Chahal HS, Wu W, Ransohoff KJ, et al: Genome-wide association study identifies 14 novel risk alleles associated with basal cell carcinoma. *Nat Commun* 7:12510, 2016
30. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, et al: A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. *Nat Genet* 40:1204-1210, 2008
31. Futreal PA, Coin L, Marshall M, et al: A census of human cancer genes. *Nat Rev Cancer* 4:177-183, 2004
32. Institute of Cancer Research: JAK1 (P23458)-Overview-Molecular target synopsis. <https://cansar.icr.ac.uk/cansar/molecular-targets/P23458>
33. Dixon JR, Selvaraj S, Yue F, et al: Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376-380, 2012
34. Michailidou K, Lindström S, Dennis J, et al: Association analysis identifies 65 new breast cancer risk loci. *Nature* 551:92-94, 2017
35. Milne RL, Kuchenbaecker KB, Michailidou K, et al: Identification of ten variants associated with risk of estrogen-receptor-negative breast cancer. *Nat Genet* 49:1767-1778, 2017
36. Labreche K, Kinnersley B, Berzero G, et al: Diffuse gliomas classified by 1p/19q co-deletion, TERT promoter and IDH mutation status are associated with specific genetic risk loci. *Acta Neuropathol* 135:743-755, 2018
37. Eckel-Passow JE, Lachance DH, Molinaro AM, et al: Glioma groups based on 1p/19q, IDH, and TERT promoter mutations in tumors. *N Engl J Med* 372:2499-2508, 2015
38. Amos CI, Wu X, Broderick P, et al: Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 40:616-622, 2008