nature biomedical engineering

Supplementary information

https://doi.org/10.1038/s41551-026-01647-1

High-resolution functional mapping of androgen receptor variants

In the format provided by the authors and unedited

Supplementary Information

High-resolution functional mapping of androgen receptor variants

Contents:

Supplementary Note 1. Comparison of our functional classifications of AR with McGill Androgen Database

Only four SNVs (3.1%, 4/129; W742L, L831V, F857L, and P914V) out of the 129 CAIS variants listed in the McGill Androgen Database were classified as functional. However, two of these variants, W742L[1] and F857L,[2] were previously shown to be functional in cell-based assays, consistent with our results. Thus, we believe that the database should be updated considering our functional results. Furthermore, the remaining two CAIS variants (L831V[3] and P914R[4]) were previously classified based on analyses of only one patient per variant without cell-based assays. Thus, the evidence supporting that these two variants cause AIS is not compelling. The fraction of non-functional variants decreased in the following order: CAIS (84%, 109/129), PAIS (81%, 64/79), and MAIS (37%, 7/19).

Supplementary Note 2. Individual evaluations of the functional effects of AR variants

The mean frequencies of W742L-, L831V-, F857L-, P914R-, and K633N-containing cells either increased or remained stable at D10 compared with D0, confirming their AR functionality. By contrast, L791P and G689E variants were depleted 0.39-fold (D10 vs. D0: 5.5% vs. 14%) and 0.23-fold (3.7% vs. 16%), respectively, consistent with their non-functional classification (Extended Data Fig. 4b). Notably, these individually measured fold changes strongly correlated with functional scores from our high-throughput assay ($r = 0.88$; Extended Data Fig. 4c), supporting the robustness of our large-scale evaluation.

Supplementary Note 3. Individual evaluation of resistance to enzalutamide and bavdegalutamide

Frequencies of L702H- and Y740C-containing cells increased 1.3-fold (enzalutamide, D10 vs. untreated, D10: 33% vs. 25%) and 1.5-fold (16% vs. 11%), respectively, under enzalutamide, and 1.4-fold (35% vs. 25%) and 1.5-fold (16% vs. 11%) under bavdegalutamide, confirming their resistance to both drugs. Similarly, M750V and F877L frequencies increased 1.9-fold (17% vs. 8.9%) and 1.8-fold (25% vs. 14%), respectively, under enzalutamide but decreased 0.79-fold (7.0% vs. 8.9%) and 0.72-fold (10.1% vs. 14%), respectively, under bavdegalutamide, supporting the recommendation of bavdegalutamide for these variants. In contrast, R630Q was depleted 0.70-fold (33% vs. 47%) and 0.60-fold (28% vs. 47%) under enzalutamide and bavdegalutamide conditions, respectively, confirming its sensitivity to both drugs (Fig. 5b).

Supplementary Note 4. Deep learning-based prediction of functional and drug-resistant effects of AR variants

Enzalutamide inhibits AR by competing with androgens for binding to the LBD, machine learning based on the AR structure may enable relatively accurate prediction of enzalutamide activity. However, bavdegalutamide (a PROTAC) acts through a more complex process that requires AR binding, ternary complex formation with an E3 ligase, and subsequent ubiquitin-mediated degradation.[5-7] This multistep mechanism introduces

multiple layers of biological variability, and PROTAC-mediated activation rather than degradation may further weaken genotype–phenotype correlations.[8] Together, these factors likely reduce the accuracy of bavdegalutamide activity prediction, although further studies are needed to reach a definitive conclusion.

Supplementary Note 5. Construction of pegRNA libraries and cloning

The pegRNA library targeting *AR* was divided into subsets of exon-specific libraries (AR-exon4 through AR-exon8, Supplementary Table1). Pooled 274-bp oligonucleotides for plasmid construction were array-synthesized by Twist Bioscience. Oligonucleotides were designed to include the following elements: (1) a 19-bp homology arm with a 3' terminus from a U6 promoter; (2) a 20-bp sequence with a G at the 5' terminus, followed by a 19-bp spacer sequence; (3) an improved version of the SpCas9 single guide RNA (sgRNA) scaffold sequence[9]; (4) a RT template and a PBS; (5) an 8-bp linker sequence obtained from pegLIT tools followed by a 37-bp tevopreQ1 and a 7-bp poly-T sequence[10]; (6) an 18-bp unique barcode sequence corresponding to a unique pegRNA sequence (barcode sequences that differed from each other by an edit distance of at least three nucleotides and that contained the constant "GTCAG" sequence for subsequent analysis were used); (7) a random "Buffer" sequence of various lengths added to pegRNAs with shorter RT templates and PBSs to ensure that all oligonucleotides were of equal length; and (8) a 20-bp unique homology arm for subset-library amplification.

Unique homology arm sequences for *AR* were as follows:

AR exon 4: 5'-TGCGACCGTAATCAAACCAA-3'
AR exon 5: 5'-GTTCAAATTGCGTGCGACAT-3'
AR exon 6: 5'-AAATGGATGCCTTGTGCGAA-3'
AR exon 7: 5'-AAACGGAGCCATGAGTTTGT-3'
AR exon 8: 5'-ACCGCGCTCGAAGAATTTAA-3'

Oligonucleotide pools encoding pegRNA sequences for each exon were PCR-amplified using Q5 High-Fidelity DNA Polymerase (NEB). PCR cycling conditions were as follows: 30 s at 98℃, 30 s at 68℃, 2 min at 72℃, for 14 cycles. The resulting amplicons were then gel-purified. To prepare the plasmid backbone, Lenti-gRNA-Puro (Addgene, 84752) was digested with BsmBI (Enzynomics) at 55°C for 6 h and subsequently gel-purified. The linearized plasmid was then assembled with the amplified oligonucleotide pool using NEBuilder® HiFi DNA Assembly Master Mix. The assembled product was concentrated via isopropanol precipitation using GlycoBlue™ Coprecipitant (Invitrogen™) and subsequently electroporated into Endura electrocompetent cells (Lucigen) using a MicroPulser (Bio-Rad) to generate a plasmid library.

Supplementary Note 6. Construction of plasmid vectors

All constructs were generated using Gibson assembly (NEBuilder® HiFi DNA Assembly Master Mix, NEB) or ligation following restriction digestion (T4 DNA Ligase, NEB) according to the manufacturer's protocol.

To construct pLenti-PEmax-BSD and pLenti-PE7-BSD, the SpCas9 sequence from the LentiCas9-Blast plasmid (Addgene, 52962) was replaced with either the PEmax or PE7 sequence that had been amplified from pCMV-PEmax (Addgene, 174820) or pCMV-PE7 (Addgene, 214812), respectively.

To construct plasmids encoding epegRNAs to be used in transfection and individual ARSI resistance evaluation experiments, the Lenti-gRNA-Puro backbone plasmid (Addgene, 84752) was digested with BsmBI-v2 (NEB), gel purified, and ligated with inserts encoding epegRNA sequences using T4 DNA ligase. The inserts were prepared by annealing multiple single-stranded DNAs (Supplementary Table4), forming double-stranded DNAs with 4-bp overhangs, which served as substrates for T4 ligation.

For bacterial transformation, 2–4 µl of either Gibson assembly reaction or T4 ligation mix was introduced into NEB Stable Competent E. coli (NEB). Cells were grown at 30°C or 37°C, and plasmid DNA was isolated using the NICSROprep™ Plasmid DNA Miniprep S&V Kit (Bionics). The resulting plasmids were confirmed by Sanger sequencing (Macrogen).

Supplementary Note 7. Lentivirus production

HEK293T cells were seeded at a density of 1×10^7 cells per 150-mm dish 18 hours prior to transfection. Lentivirus was generated using Lipofectamine 3000 (Thermo Fischer Scientific) following the manufacturer's protocol. Briefly, 20 µg of transfer plasmid containing the gene of interest, 15 µg of psPAX2 (Addgene #12260), and 5 µg of pMD2.G (Addgene #12259) were combined and diluted in 1 mL of Opti-MEM (Gibco), followed by the addition of 80 µL of P3000 reagent. Separately, 90 µL of Lipofectamine 3000 was diluted in 1 mL of Opti-MEM and then mixed with the plasmid mixture. After incubating for 15 min, the mixture was added to cells. The culture medium was refreshed after 6 hours, and virus-containing supernatant was collected 48 hours post-transfection. The supernatant was filtered using a 0.45 µM Millex-HV low protein-binding membrane (Millipore). The filtered virus was aliquoted and stored at −80°C for future use.

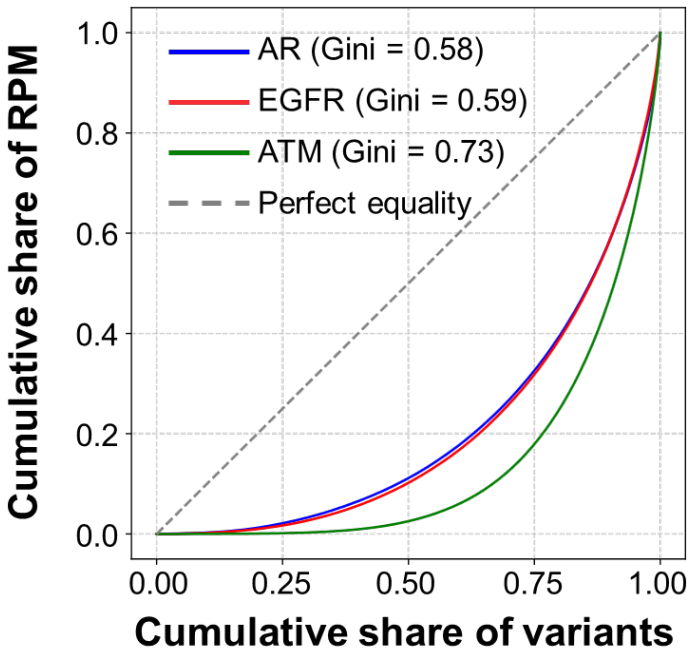
Supplementary Note 8. Genomic DNA extraction and deep sequencing

gDNA was extracted using the Wizard® Genomic DNA Purification Kit (Promega) according to the manufacturer's protocol. PCR amplification was performed in two rounds.

In the first round, 72 to 144 µg of purified gDNA (ensuring >5,000× coverage per library, assuming 6 pg of gDNA per cell) was amplified using exon-specific primers and

Q5® High-Fidelity DNA Polymerase (NEB) (Supplementary Table5). PCR conditions were set with an annealing temperature range of 57–71°C and 24 cycles. The resulting PCR products were pooled, purified using a MEGAquick-Spin Total Fragment DNA Purification kit (iNtRON Biotechnology), and further size-selected via 2% agarose gel electrophoresis and gel extraction using the same purification kit.

For indexing PCR, 40 ng of purified first-round PCR product was used as a template and amplified with Illumina indexing primers using Q5® High-Fidelity DNA Polymerase. The reaction was conducted at an annealing temperature of 60°C for 8 cycles. The final PCR products were purified using the same purification kit and subsequently sequenced on a NovaSeq 6000 system (Illumina).



Supplementary Fig. 1 Assessment of AR SAAV library homogeneity. Lorenz curve analysis of normalized read counts (RPM) across all single amino acid variants (SAAVs) 10 days after transduction. The cumulative proportion of variants (x-axis) was plotted against the cumulative proportion of sequencing reads (y-axis). Perfectly uniform representation would follow the diagonal (dashed line), whereas deviations indicate unequal distribution. The observed Lorenz curve (blue) demonstrated that all intended variants were represented with high confidence. The calculated Gini coefficient (0.58) indicates moderate variation in relative abundance across variants, which is expected given sequence-dependent differences in prime editing efficiency, but confirms that no variants were lost from the library.

Variant	Target	Region	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Wild-type (%)	Intended-edit (%)	Indel (%)
W742L	On	chrX: 67717508	G	G	C	T	G	T	C	A	T	T	C	A	G	T	A	C	T	C	C	T	G	G	78.1	21.8	0.1
	Off_1	chr8: 96354944	G	G	C	T	G	T	C	A	T	T	C	A	c	a	A	C	T	C	C	A	G	G	99.7	0	0.3
	Off_2	chr15: 63110704	G	G	C	T	G	T	C	A	T	T	C	A	G	T	g	C	c	C	C	A	G	G	99.7	0	0.3
	Off_3	chr8: 74281857	G	G	C	T	G	c	C	A	T	T	C	A	G	c	t	C	T	C	C	T	G	G	99.9	0	0.1
	Off_4	chr15: 25647425	G	G	C	T	G	T	C	t	T	g	C	A	G	g	A	C	T	C	C	A	G	G	100	0	0
	Off_5	chr15: 88449499	G	G	C	T	G	g	g	A	T	T	C	A	G	g	A	C	T	C	C	T	G	G	99.8	0	0.2
L831V	On	chrX: 67722818	C	A	C	A	T	C	A	G	T	T	C	C	A	G	T	G	G	A	T	G	G	G	92.5	4.9	2.6
	Off_1	chr15: 86382134	C	A	C	A	T	C	A	G	T	c	C	C	A	G	g	G	G	A	T	G	G	G	100	0	0
	Off_2	chr8: 74721174	C	A	C	c	T	g	t	G	T	T	C	C	A	G	T	G	G	A	T	G	G	G	99.9	0	0.1
	Off_3	chr22: 24899357	C	A	C	A	T	C	A	G	g	T	g	C	A	G	T	G	G	A	g	T	G	G	100	0	0
	Off_4	chr7: 51877534	C	A	C	A	T	C	c	t	T	T	C	C	A	G	T	G	G	g	T	T	G	G	99.7	0	0.3
	Off_5	chr15: 33360725	C	A	G	C	C	A	G	T	G	T	G	T	t	g	G	A	g	T	G	G	G	G	99.7	0	0.3
F857L	On	chrX: 67722950	G	T	C	C	A	G	G	A	G	C	T	T	G	G	T	G	A	G	C	T	G	G	74.8	25.1	0.1
	Off_1	chr7: 84906520	G	T	C	C	A	t	G	A	G	t	T	T	G	G	T	G	A	G	C	A	G	G	99.8	0	0.2
	Off_2	chr2: 111197176	a	T	a	C	A	G	G	A	G	C	T	T	G	G	T	G	A	G	C	T	G	G	99.7	0	0.3
	Off_3	chr14: 74584778	G	a	C	C	A	G	G	A	t	C	T	T	G	G	T	G	A	G	C	T	G	G	99.4	0	0.6
	Off_4	chr8: 41270173	G	T	C	C	A	G	a	A	G	C	T	T	G	G	T	G	A	t	a	T	G	G	99.8	0	0.2
	Off_5	chr8: 128581767	G	T	C	C	A	a	G	A	G	C	T	T	a	G	a	G	A	G	C	A	G	G	99.6	0	0.4
L791P	On	chrX: 67721859	C	A	G	C	C	A	G	T	G	T	G	T	C	C	G	A	A	T	G	A	G	G	84.6	14.2	1.2
	Off_1	chr16: 49538482	C	A	G	g	C	A	G	T	G	T	G	T	C	t	G	A	A	T	G	T	G	G	99.6	0	0.4
	Off_2	chr9: 130416720	C	A	G	C	C	A	G	T	G	T	G	T	C	C	t	A	A	a	G	T	G	G	99.3	0	0.7
	Off_3	chr2: 146370853	C	A	G	C	C	t	G	T	G	T	c	T	C	t	G	A	A	T	G	T	G	G	100	0	0
	Off_4	chr8: 142009502	C	A	G	C	C	A	G	T	G	T	G	c	C	C	c	A	A	g	G	A	G	G	99.7	0	0.3
	Off_5	chr2: 146370853	C	A	G	C	C	t	G	T	G	T	c	T	C	t	G	A	A	T	G	T	G	G	100	0	0

Supplementary Fig. 2 Evaluation of off-target effects. DNA sequences at on-target and potential off-target sites were analyzed by deep sequencing in four variant clones generated with the pegRNAs used for individual functional assays. Numbers at the top indicate positions within the protospacer (1–19) and protospacer-adjacent motif (NGG PAM, 20–22, grey). Base mismatches between on- and off-target sites are highlighted in orange. Frequencies of wild-type sequences, intended edits, and indels were determined by deep sequencing.

References

1. Lallous, N. *et al.* Functional analysis of androgen receptor mutations that confer anti-androgen resistance identified in circulating cell-free DNA from prostate cancer patients. *Genome Biol* **17**, 10 (2016).
2. Mongan, N. P. *et al.* Two de novo mutations in the AR gene cause the complete androgen insensitivity syndrome in a pair of monozygotic twins. *J Clin Endocrinol Metab* **87**, 1057-1061 (2002).
3. Chavez, B., Vilchis, F., Zenteno, J. C., Larrea, F. & Kofman-Alfaro, S. Novel molecular defects in the androgen receptor gene of Mexican patients with androgen insensitivity. *Clin Genet* **59**, 185-188 (2001).
4. Wu, X., Zhou, Q., Mao, J., Lu, S. & Nie, M. Mutational analysis of androgen receptor gene in four Chinese patients with male pseudohermaphroditism. *Fertil Steril* **93**, 2076 e2071-2074 (2010).
5. Burke, M. R., Smith, A. R. & Zheng, G. Overcoming Cancer Drug Resistance Utilizing PROTAC Technology. *Front Cell Dev Biol* **10**, 872729 (2022).
6. Rutherford, K. A. & McManus, K. J. PROTACs: Current and Future Potential as a Precision Medicine Strategy to Combat Cancer. *Mol Cancer Ther* **23**, 454-463 (2024).
7. Snyder, L. B. *et al.* Preclinical Evaluation of Bavdegalutamide (ARV-110), a Novel PROteolysis TArgeting Chimera Androgen Receptor Degrader. *Mol Cancer Ther* (2024).
8. Huber, A. D., Lin, W., Poudel, S., Miller, D. J. & Chen, T. PROTAC-mediated activation, rather than degradation, of a nuclear receptor reveals complex ligand-receptor interaction network. *Structure* **32**, 2352-2363 e2358 (2024).
9. Dang, Y. *et al.* Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biol* **16**, 280 (2015).
10. Nelson, J. W. *et al.* Engineered pegRNAs improve prime editing efficiency. *Nat Biotechnol* **40**, 402-410 (2022).