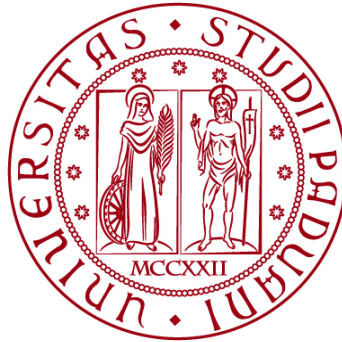


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea in Biologia Molecolare



ELABORATO DI LAUREA

**MODIFICAZIONE DEL GENOMA MITOCONDRIALE
ATTRAVERSO L'EDITOR DI BASI DI CITOSINA (DdCBE)
MEDIANTE VIRUS ADENO-ASSOCIATI (AAV) IN TOPI
ADULTI E NEONATI**

**Tutor: Prof. Carlo Fiore Viscomi
Dipartimento di Scienze Biomediche**

Laureanda: Arianna Breda

ANNO ACCADEMICO 2022/2023

INDICE

1. ABSTRACT	1
2. INTRODUZIONE	3
3. DdCBE E TRASMISSIONE IN VIVO	7
3.1 Trasmissione in vivo di DdCBE in topi adulti	10
3.2 Trasmissione in vivo di DdCBE in topi neonati	11
4. METODI SPERIMENTALI	13
4.1 Costruzione di plasmidi TALE e di vettori virali	13
4.2 Estrazione genoma mitocondriale.....	15
4.3 High-throughput sequencing	15
4.4 Real-time PCR.....	16
5. DISCUSSIONE E CONCLUSIONE	17
6. BIBLIOGRAFIA.....	19
APPENDICE	

1. ABSTRACT

I mitocondri sono organelli citoplasmatici delle cellule eucariotiche nei quali avviene il processo della respirazione cellulare che fornisce energia, sotto forma di ATP, alle cellule stesse. Sono dotati di un piccolo genoma (mtDNA) codificante una parte delle proteine mitocondriali, le restanti derivano dal genoma nucleare. Mutazioni nei geni, sia nucleari sia mitocondriali codificanti le proteine mitocondriali, alterano la funzionalità del mitocondrio innescando patologie mitocondriali.

Per molto tempo, i ricercatori cercarono di studiare le basi biologiche di tali patologie riscontrando, però, diverse problematiche come la limitata disponibilità di strumenti e l'inefficiente trasporto di acidi nucleici all'interno degli organelli stessi.

Gli autori dell'articolo hanno sviluppato una nuova tecnica per modificare il genoma mitocondriale. Si tratta dell'editor di basi di citosina (DdCBE) derivato dalla DNA-deaminasi a doppio filamento (DddA). Gli esperimenti sono stati eseguiti in topi vivi utilizzando virus adeno-associati (AAV) come vettori per la trasmissione di DdCBE nel tessuto cardiaco. I dati raccolti confermano che DdCBE è in grado di introdurre mutazioni nei geni d'interesse nei topi adulti e neonati ma con un'efficienza maggiore in questi ultimi.

In futuro DdCBE potrebbe essere utilizzato nell'ambito sia della ricerca sia terapeutico, correggendo mutazioni puntiformi patogeniche.

2. INTRODUZIONE

I mitocondri sono organelli localizzati nel citoplasma di tutti i tipi cellulari dell'organismo umano, ad eccezione degli eritrociti. Questi organelli hanno una forma allungata e presentano una dimensione compresa tra 0,3 e 3 μm . Sono delimitati da due membrane, una interna ed una esterna. La regione tra le due membrane è definita spazio intermembrana mentre quella delimitata dalla membrana interna, matrice mitocondriale.

Le due membrane presentano una composizione e una funzione differente. Quella esterna è formata in ugual quantità da fosfolipidi e proteine, delle quali la maggior parte sono porine. La loro presenza determina la permeabilità della membrana esterna a piccole molecole aventi una dimensione inferiore di 10KDa. La membrana interna è costituita da 70% di proteine e 30% di lipidi ed è fondamentalmente impermeabile ai soluti. Tuttavia, la presenza di numerosi trasportatori media il trasporto di diversi metaboliti quali piruvato, nucleotidi, ADP ect... Inoltre, la membrana interna è caratterizzata dalle "creste mitocondriali" ossia ripiegamenti della membrana stessa. Tali strutture hanno il compito di aumentare la superficie in quanto contengono i complessi proteici coinvolti nella fosforilazione ossidativa. Nella matrice mitocondriale si trova il DNA mitocondriale (mtDNA) e numerosi ribosomi ed enzimi.

I mitocondri svolgono dei ruoli fondamentali per le singole cellule. Sono coinvolti principalmente nella respirazione cellulare ma anche in diverse vie metaboliche come l'ossidazione degli acidi grassi e la gluconeogenesi. Inoltre, partecipano anche in processi cellulari come apoptosi, deposito di ioni calcio e sintesi del gruppo eme.

I mitocondri sono considerati la centrale energetica della cellula in quanto al loro interno avviene la respirazione cellulare. Il metabolismo energetico comprende varie vie metaboliche, che convergono verso il ciclo di Krebs e la catena di trasporto degli elettroni, che in ultima analisi supporta il processo della fosforilazione ossidativa. Queste ultime due avvengono all'interno del mitocondrio ma, in particolare, questa tesi si concentrerà sugli effetti delle mutazioni nei geni che codificano i complessi della catena di trasporto elettronico e di sintesi dell'ATP.

La catena respiratoria comporta il trasferimento di elettroni che derivano dal metabolismo intermedio di zuccheri, lipidi e proteine in forma di equivalenti di riduzione (NADH e FADH₂) all'ossigeno. Questo trasporto coinvolge quattro complessi proteici, denominati complesso I, II, III e IV, e due trasportatori mobili, quali ubiquinone e citocromo c.

Il complesso I accetta due elettroni dal NADH e li trasferisce all'ubichinone, o coenzima Q, che viene ridotto a ubichinolo. Quest'ultimo è un trasportatore mobile, ossia una proteina in grado di muoversi all'interno della membrana interna, che riceve elettroni non solo dal complesso I ma anche dal complesso II e

li trasferisce al complesso III. Il complesso II riceve gli elettroni dal FADH₂ e, analogamente al complesso I, è in grado di ridurre l'ubichinone a ubichinolo. Il complesso III catalizza l'ossidazione dell'ubichinolo a ubichinone, trasferendo gli elettroni al citocromo c; secondo trasportatore mobile della catena respiratoria. Quest'ultimo, a differenza dell'ubichinone, è adeso alla membrana interna sul lato intermembrana. Il citocromo c ridotto viene ossidato dal Complesso IV, che trasferisce gli elettroni all'ossigeno molecolare riducendolo ad acqua.

L'energia che deriva dal trasporto di elettroni viene utilizzata dai complessi per pompare protoni dalla matrice mitocondriale verso lo spazio intermembrana. In particolare, i complessi I e III trasferiscono 4 protoni, il complesso IV due mentre il complesso II non trasferisce nessun protone. Considerando l'impermeabilità della membrana interna ai protoni, questo genera un gradiente protonico che verrà sfruttato dall'ATP sintasi per sintetizzare ATP a partire dall'ADP e P inorganico.

Il mtDNA ha una dimensione di circa 16,6Kb ed è una molecola di DNA circolare a doppio filamento. In base al contenuto di guanina, i due filamenti vengono distinti in "catena leggera" e "catena pesante". Il mtDNA è formato da 37 geni di cui 13 codificano proteine coinvolte nella fosforilazione ossidativa, 22 tRNA e 2 rRNA. Il proteoma mitocondriale è rappresentato da circa 1500 proteine totali; quindi, solo un 10% di proteine viene codificato direttamente dal mitocondrio. Le restanti derivano da geni del genoma nucleare (nDNA), tradotte nel citosol e poi importate nel mitocondrio.

Il mtDNA presenta delle caratteristiche differenti dal nDNA. Innanzitutto, il mtDNA si replica indipendentemente dal nDNA e durante tutte le fasi del ciclo cellulare. Infatti, un mitocondrio non possiede un singolo genoma ma bensì centinaia di copie. Il nDNA possiede molti promotori, ognuno localizzato a monte del gene che controllano, mentre il mtDNA possiede tre promotori totali:

- HSP1, si trova nel filamento pesante e permette la trascrizione dei due rRNA;
- HSP2, è localizzato a monte di HSP1 e permette la trascrizione del resto del filamento pesante in un unico trascritto policistronico;
- LPS, è collocato nel filamento leggero garantendone la sua trascrizione.

Il filamento pesante contiene anche una regione non codificante di 1,1kb denominata D-loop. Quest'ultimo ha il ruolo di controllare la replicazione e la trascrizione del genoma stesso.

Anche a livello del codice genetico ci sono alcune differenze tra mtDNA e nDNA, infatti, ci sono alcuni codoni che hanno un significato diverso tra i due genomi. Ad esempio, UGA codifica il triptofano nel mtDNA mentre nel nDNA rappresenta un codone di stop.

Il tasso di mutazione del mtDNA è maggiore rispetto a quello del nDNA. Le cause principali sono l'assenza di istoni e la presenza di specie reattive dell'ossigeno (ROS). Il primo rende il mtDNA meno compatto e di conseguenza più suscettibile all'azione dei ROS, prodotti durante la fosforilazione ossidativa. Inoltre, a livello

del mtDNA non agiscono i sistemi di riparazione del DNA nucleare e questo favorisce l'accumulo di mutazioni. Tutto questo fa sì che il tasso di mutazione del mtDNA sia 20 volte superiore rispetto a quello del nDNA. È stato dimostrato che in tutti gli individui, molecole di mtDNA wild-tipe e mutate possono coesistere.

Si parla di omoplasma quando all'interno di una cellula tutti i mitocondri presentano lo stesso genoma. Si parla di eteroplasma quando vi è la presenza contemporanea di mtDNA wild-tipe e mtDNA mutato all'interno della stessa cellula. Il livello di eteroplasma può variare, non solo tra cellule dello stesso tessuto, ma anche tra tessuti differenti dello stesso individuo e anche tra individui. Le mutazioni eteroplasmiche possono determinare lo sviluppo di patologie mitocondriali se, la proporzione di mtDNA mutato, supera una certa soglia che si aggira tra il 60-90% a seconda della mutazione (effetto soglia). Tale valore viene definito come la percentuale di mitocondri mutati, sulla numerosità totale, sufficiente per provocare il fenotipo patologico. Questo significa che la presenza di una quantità di mitocondri mutati al di sotto della soglia critica non porta alla manifestazione della patologia in quanto vengono compensati da quelli wild-tipe. Inoltre, è la percentuale di mitocondri mutati che determina il fenotipo clinico.

Il livello di eteroplasma può essere influenzato da almeno tre fattori. Il primo è la "segregazione vegetativa" secondo cui la ripartizione degli organelli tra le cellule figlie può essere asimmetrica. Di conseguenza, una cellula figlia può ricevere una quantità di mitocondri mutati maggiore o minore rispetto alla seconda.

Il secondo fattore è la "replicazione rilassata" nelle cellule postmitotiche. Nel caso della replicazione del mtDNA, non esiste nessun meccanismo che controlla quante volte o quale molecola viene replicata. Quindi, se viene replicata con maggior frequenza la molecola di mtDNA mutato rispetto al wild-tipe questo induce un aumento del livello di eteroplasma, aumentando la probabilità che essa diventi significativa.

Il terzo fattore è conosciuto come "collo di bottiglia genetico" e si verifica durante lo sviluppo degli oociti. In particolare, gli oociti attraversano una fase nella quale subiscono una riduzione della quantità di mtDNA, che viene poi ristabilita durante la maturazione. Perciò, potranno essere ripopolati da mtDNA wild-tipe o mutato. Il fenomeno del collo di bottiglia è il responsabile dell'eteroplasma variabile tra i diversi figli della stessa madre, dunque, della variabilità dei sintomi anche in membri della stessa famiglia. Infatti, una madre affetta da una patologia mitocondriale non è detto che abbia tutti figli che manifestano la patologia o che i sintomi abbiano tutti lo stesso livello di gravità.

Le malattie mitocondriali sono un insieme di patologie causate da mutazioni in geni nucleari o mitocondriali codificanti le proteine coinvolte nella fosforilazione ossidativa. Questo genera mitocondri che presentano dei difetti nella

respirazione cellulare e nella sintesi di ATP. Dato che il mtDNA presenta un'eredità di tipo uni-parentale, o materna, le patologie dovute alle mutazioni in geni mitocondriali verranno trasmesse dalla madre affetta a tutti i figli di entrambi i sessi. Inoltre, tali patologie possono insorgere sia da condizioni omoplasmiche sia eteroplasmiche. Invece, quelle causate da mutazioni nei geni nucleari vengono trasmesse secondo le regole dell'eredità mendeliana.

In generale, le malattie mitocondriali possono insorgere a qualsiasi età e con una grande variabilità anche a livello di sintomatologia, a causa del livello di eteroplasmia. Possono colpire qualsiasi organo o tessuto, assumendo quindi fenotipi clinici differenti.

La diagnosi è difficile a causa della debole correlazione tra fenotipo e genotipo. Infatti, una mutazione è in grado di originare diversi fenotipi clinici ma anche la stessa patologia può essere innescata da mutazioni diverse. La diagnosi viene ulteriormente complicata dal fatto che i sintomi possono variare anche a seconda dello stile di vita, del sesso, condizioni ambientali ed epigenetiche.

È per questi motivi che la diagnosi avviene principalmente con l'individualizzazione della mutazione attraverso il sequenziamento del mtDNA e/o nDNA. Viene anche utilizzata la biopsia muscolare per effettuare un'analisi istochimica e biochimica del tessuto. Attualmente non esistono cure per le malattie mitocondriali e i trattamenti si limitano ad alleviare i sintomi.

Lo studio delle malattie mitocondriali è da sempre considerato molto complicato. Infatti, per molto tempo, la ricerca è stata condotta solo a carico di cellule derivanti direttamente dai pazienti. Questo perché il mtDNA è difficilmente modificabile, aumentando la difficoltà di creare modelli rappresentativi di una determinata patologia.

A livello sperimentale, negli ultimi 15 anni sono stati utilizzati enzimi di restrizione indirizzati ai mitocondri e in grado di riconoscere sequenze nucleotidiche specifiche del mtDNA e di tagliarle. In questo modo le molecole di mtDNA mutate venivano linearizzate e degradate riducendo il livello di eteroplasmia al di sotto della soglia critica. I limiti di questo approccio sono (1) può essere utilizzato solo in presenza della specifica sequenza riconosciuta dall'endonucleasi scelta (2) non consente di modificare il mtDNA in modo specifico. Anche il sistema CRISPR/Cas è inutilizzabile per modificare il mtDNA in quanto gli RNA guida non vengono importati nei mitocondri.

Recentemente, è emerso un nuovo strumento in grado di modificare il mtDNA in modo sito-specifico. Si tratta dell'editor di basi di citosina (DdCBE) che è in grado di convertire le citosine in uracile.

3. DdcBE E TRASMISSIONE IN VIVO

L'editor di basi di citosina DdcBE (DddA-derived cytosine base editor) è stato ottenuto a partire da una DNA deaminasi batterica. Si tratta di DddA, una tossina prodotta dal batterio *Burkholderia cenocepacia* e secreta nell'ambiente esterno come antibatterico in situazioni di competizione. Tale enzima catalizza, all'interno delle molecole a DNA a doppio filamento (double strand DNA, dsDNA), la rimozione del gruppo amminico della citosina con la formazione dell'uracile. Quest'ultima non è una base azotata tipica del DNA ma si comporta come un'altra base normalmente presente nel DNA, la timina. Dunque, quando questo sito modificato viene replicato, l'apparato di replicazione interpreta l'uracile come una timina introducendo l'adenina nel filamento in sintesi. Nel successivo ciclo di replicazione l'adenina verrà appaiata alla timina. In questo modo avviene la conversione della citosina in timina con la transizione del sito da CG a TA. Questa attività aspecifica rende DddA tossica per le cellule in quanto potrebbe convertire tutte le citosine, di un dsDNA, in uracile. Per questo motivo gli autori dell'articolo hanno diviso l'enzima in due parti, così da renderlo inattivo fino a quando le due non vengono riunite nel DNA bersaglio.

L'attività catalitica di DddA è svolta dal dominio denominato DddAtox, costituito da alfa eliche e da 5 foglietti beta. DddA è stata divisa in modo tale che una porzione contenga N-terminale del dominio DddAtox e l'altra il C-terminale. Di conseguenza, l'enzima sarà attivo solo quando N-terminale e C-terminale saranno vicini tra loro, formando così il sito catalitico completo.

In particolare, il sito catalitico è stato diviso in due modi identificati come G1333 e G1397 per ognuno dei quali vengono ottenute due porzioni: N-terminale e C-terminale. Ad esempio, per G1333 si parlerà rispettivamente di G1333-N e G1333-C. Per garantire l'azione di DddAtox solo all'interno delle regioni d'interesse, ogni metà è stata fusa con i domini TALE.

Gli effettori TAL (Transcription Activator Like), o proteine TALE, sono prodotte da proteobatteri, la maggior parte appartenenti al genere *Xanthomonas*, che infettano le piante. Tali microrganismi iniettano gli effettori TAL all'interno della cellula vegetale portando, in essa, l'attivazione di geni che rendono la pianta suscettibile all'infezione.

TALE sono proteine simili ai fattori di trascrizione, infatti, sono costituite da domini in grado di riconoscere e di legare una sequenza specifica del DNA.

La caratteristica principale del dominio TALE è la composizione in sequenze, di 33-34 residui amminoacidici, ripetute in tandem. Ogni ripetizione rappresenta un modulo del dominio che è specifico per una determinata base del DNA. Questo è stato confermato dall'osservazione, da parte dei ricercatori, che il numero di ripetizioni coincide con la lunghezza della sequenza target.

La maggior parte degli amminoacidi di un modulo sono conservati ad eccezione dei residui in posizione 12 e 13 che risultano iper-variabili. Tali residui definiscono

qual è la base del DNA che verrà riconosciuta da quel determinato modulo.

I TALE sono domini molto utilizzati nel campo dell'ingegneria genetica perché possono essere fusi con enzimi non specifici, determinando così la loro azione solo a livello di sequenze specifiche.

Gli autori dell'articolo hanno realizzato quattro coppie di DdCBE, ognuna costituita da un editor specifico per il filamento leggero (DdCBE-L) del mtDNA e uno per il filamento pesante (DdCBE-H). Dunque, per i motivi spiegati sopra, DdCBE è in grado di catalizzare la deaminazione della citosina solo quando DdCBE-L e DdCBE-H si trovano legati alla sequenza target dei rispettivi filamenti.

Considerando l'architettura di DdCBE-L (**Figura 1**), essa comprende: sequenza MTS (mitochondrial targeting sequence), dominio N-terminale, dominio TALE, dominio C-terminale, metà di DddAtox che può derivare dalla divisione G1333 o G1397, e proteina UGI.

La sequenza MTS garantisce l'indirizzamento e l'import dell'editor nei mitocondri. Si tratta di una specifica sequenza amminoacidica, ricca di amminoacidi carichi positivamente, che viene riconosciuta dai trasportatori localizzati sulla membrana del mitocondrio, garantendone così l'import nella matrice mitocondriale.

Il dominio TALE è costituito da moduli che riconoscono la sequenza d'interesse localizzata nel filamento leggero. Bisogna considerare che maggiore è la numerosità dei moduli e maggiore sarà la specificità, in quanto maggiore è la lunghezza della sequenza e minore è la probabilità di trovare una sequenza identica per caso e quindi aspecifica. In questo studio, i domini TALE sono stati progettati per riconoscere una sequenza di 16 bp. I domini N e C terminali, fiancheggiando il dominio TALE, derivano dalla proteina TALE di origine e sono necessari per favorire il legame del dominio TALE al DNA in quanto creano dei contatti con le basi azotate.

UGI è un inibitore dell'enzima uracil-DNA glicosilasi mitocondriale (UGI1). Quest'ultimo catalizza la rimozione dell'uracile, originatosi dalla deaminazione della citosina, diminuendo così la comparsa di questo tipo di mutazione. Questo è supportato da precedenti studi dove è stato dimostrato che la presenza di UGI nel costrutto aumenta il numero di residui di uracile rilevati, in seguito all'azione dell'editor, rispetto ad uno privo di UGI.

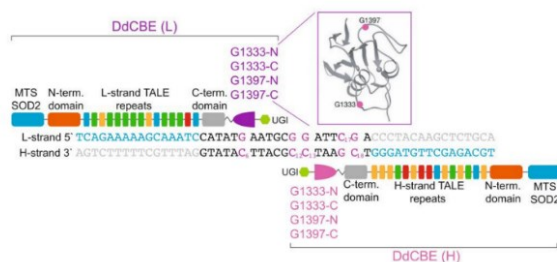


Figura 1: sono rappresentati i due monomeri di DdCBE ognuno legato al filamento corrispondente. Per ogni monomero la porzione azzurra rappresenta la sequenza MTS, arancione il dominio N-terminale, ogni modulo del dominio TALE è associato a un colore diverso in base

alla base azotata che riconosce (azzurro citosina, verde adenina, giallo guanina e rosso timina), grigio il dominio C-terminale, viola o rosa una metà di DddAtox e verde la

proteina UGI. Entrambi i domini TALE sono stati progettati per riconoscere la sequenza di 16bp evidenziata in azzurro. La sequenza, all'interno della quale DdCBE è in grado di agire, è di 19bp ma le citosine di interesse sono solo C12 e C13 nel filamento pesante 3'-5'.

In particolare, gli autori hanno creato tali coppie per indurre mutazioni nel codone della glicina 40 localizzato all'interno di *MT-ND3*. Tale gene codifica la subunità 3 del complesso I ed è ortologo al gene *MT-ND3* umano. Quest'ultimo è coinvolto in numerose patologie mitocondriali, tra le quali la neuropatia ottica ereditaria di Leber (LHON) e la sindrome di Leigh (LS). LHON è una patologia neurodegenerativa del nervo ottico che provoca la perdita bilaterale della vista. In particolare, porta alla morte delle cellule gangliari della retina; tipologia cellulare che costituisce il nervo ottico e trasporta le informazioni visive dalla retina alla corteccia cerebrale. Si tratta di una malattia rara, infatti è stata stimata un'incidenza pari a 1/100.000, con un esordio durante l'adolescenza o in età adulta. LS è una patologia neurologica che colpisce il sistema nervoso centrale provocando lesioni al tronco encefalico e ai gangli della base dell'encefalo. Si manifesta durante la prima infanzia con acidosi lattica, interruzione dello sviluppo psicomotorio, convulsioni, paralisi extraoculare e perdita della forza muscolare. Anch'essa è rara, infatti, colpisce 1-9 individui ogni 100.000.

Il codone 40 della glicina è rappresentato da GGA, quindi, sono state mirate le citosine ad esso complementare ossia C12 e C13. Come si può notare dalla figura 1, le due citosine di interesse sono precedute da una timina. Uno studio precedente ha dimostrato che, se due citosine consecutive sono precedute da una timina, esse possono andare incontro a tre modificazioni differenti (1) entrambe le citosine subiscono la deaminazione mutando, così, il codone da glicina a lisina (G40K), (2) solo C13 viene deaminata, originando un codone codificante per l'acido glutammico (G40E), (3) la deaminazione a carico solo di C12 porta alla formazione di un codone di stop prematuro (G40*).

All'inizio di questo studio, gli autori hanno testato tutte e quattro le coppie di DdCBE con lo scopo di individuare quella in grado di introdurre un maggior numero di modificazioni. Per fare ciò è stato eseguito un esperimento in vitro secondo il quale sono stati co-espressi plasmidi, contenenti i monomeri di DdCBE, in colture cellulari. A partire da queste cellule sono state quantificate, attraverso analisi di sequenziamento di ultima generazione (NGS), il numero di citosine modificate in relazione alle tre mutazioni precedentemente descritte.

La coppia 1 è costituita da DdCBE-L contenente la porzione C-terminale G1333 del dominio DddAtox e da DdCBE-H contenente la porzione N-terminale di G1333. Per tale coppia è stata rilevata la formazione della mutazione G40K per il 92,5%, G40E per il 4,5% mentre G40* per il 3% (**Figura 2**). Anche le altre tre coppie sono in grado di introdurre tutte le possibili mutazioni ma con un'efficienza minore. Per questo motivo gli autori hanno scelto di utilizzare solo

la coppia 1 per gli esperimenti in vivo utilizzando virus adeno-associati (AAV) in topi adulti e neonati. Gli autori hanno eseguito l'esperimento in topi neonati in quanto il cuore presenta una dimensione minore rispetto a quella degli adulti, risultando perciò più facile da infettare. Dunque, i neonati permettono di valutare sia se i AAV rappresentano un metodo efficace per veicolare DdCBE e sia le capacità di modifica dell'editor stesso. Tuttavia, gli autori hanno testato DdCBE anche in topi adulti perché le malattie mitocondriali possono manifestarsi fin dalla nascita ma anche in tempi successivi.

I virus adeno-associati vengono usati nel campo dell'ingegneria genetica come vettori virali per esprimere geni esogeni. Questo perché sono in grado di trasferire il loro genoma all'interno della cellula ospite, ma non di provocare un'infezione senza la collaborazione di un virus Helper co-infettante. Questo è dovuto all'assenza, nel genoma virale, di elementi genici implicati nella replicazione del genoma e nell'assemblaggio dei nuovi componenti virali.

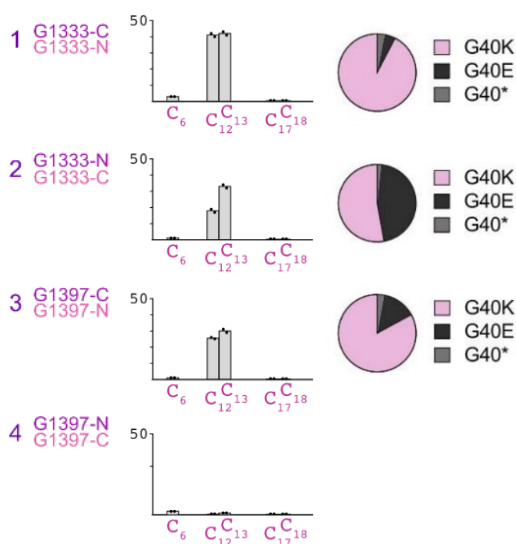


Figura 2: per ogni coppia di DdCBE, i dati ottenuti sono rappresentati attraverso un istogramma e un diagramma a torta. Il primo grafico riporta nell'asse delle ascisse le citosine contenute all'interno della sequenza target dell'editor mentre nell'asse delle ordinate il numero di modifiche rilevate a carico di esse. Il diagramma a torta illustra la porzione relativa delle tre mutazioni rispetto alla quantità totale di modifiche rilevate.

3.1 Trasmissione in vivo di DdCBE in topi adulti

Lo studio è stato eseguito in topi di 8 settimane di vita suddividendoli in tre gruppi. A quelli appartenenti al primo gruppo è stata trasmessa la coppia 1 di DdCBE, attraverso un'iniezione venosa nella coda, grazie a un ceppo di AAV specifico per il tessuto cardiaco. In particolare, è stata inoculata una miscela contenente vettori in ognuno dei quali erano presenti gli elementi genetici in grado di codificare, all'interno del citoplasma della cellula ospite, DdCBE-L o DdCBE-H. Il secondo gruppo sono i topi di controllo, nei quali non è stata trasmessa la coppia 1 ma solo il vettore virale. Al terzo gruppo appartengono i topi nei quali sono stati iniettati dei DdCBE cataliticamente inattivi. Per ognuno dei tre gruppi, alcuni topi sono stati sacrificati dopo 3 settimane dall'iniezione, i restanti dopo 24 settimane. In entrambi i casi i relativi tessuti cardiaci sono stati analizzati con NGS.

In particolare, nel tessuto cardiaco dei topi sacrificati a tre settimane dopo l'iniezione (primo gruppo) è stato osservato un editing a carico delle basi C12 e C13 di 1-2% (**Figura 3A**), mentre per i campioni relativi al secondo e terzo gruppo non è stata rilevata nessuna modifica a carico delle citosine di interesse.

Invece, nel tessuto cardiaco dei topi del primo gruppo, sacrificati dopo 24 settimane dall'iniezione, è stato osservato un editing pari a 10-20 % (**Figura 3B**) mentre per quelli relativi al secondo e terzo gruppo i risultati concordano con quelli ottenuti dai topi sacrificati dopo 3 settimane dall'iniezione.

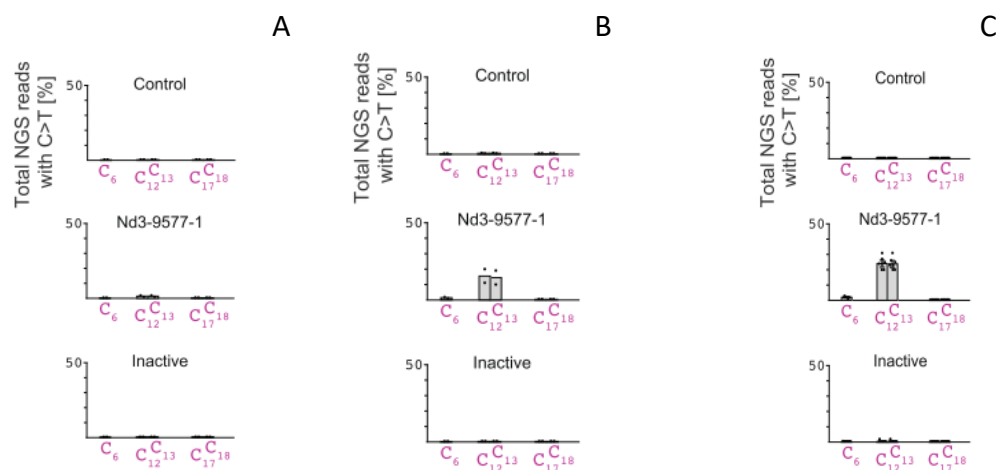


Figura 3: numero di modifiche rilevate (asse delle ordinate) a carico delle citosine presenti nella sequenza bersaglio di DdCBE (asse delle ascisse) in A: topi adulti sacrificati dopo 3 settimane; B: topi adulti sacrificati dopo 24 settimane; C: topi neonati. Gli istogrammi denominati “control” rappresentano i risultati ottenuti dai topi di controllo (secondo gruppo), quelli denominati “Nd3-9577-1” i risultati dei topi nei quali è stata iniettata la coppia 1 di DdCBE (primo gruppo), infine, quelli denominati “inactive” i risultati ottenuti dai topi nei quali è stata iniettata DdCBE cataliticamente inattiva (terzo gruppo).

3.2 Trasmissione in vivo di DdCBE in topi neonati

Lo studio è stato condotto in topi a 24 ore dalla nascita seguendo la stessa modalità precedentemente descritta per gli adulti. La differenza è che i topi poi sono stati tutti sacrificati dopo 3 settimane dall'iniezione. L'analisi del tessuto cardiaco, appartenente ai topi del primo gruppo, attraverso NGS ha portato all'osservazione di un editing pari al 20/30% (**Figura 3C**). Anche in questo caso non è stata rilevata nessuna modifica a carico delle citosine d'interesse sia nei topi neonati di controllo e sia per quelli nei quali sono stati iniettati DdCBE cataliticamente inattivi.

4. METODI SPERIMENTALI

4.1 Costruzione di plasmidi TALE e di vettori virali

Per la costruzione di plasmidi che codificano il dominio TALE possono essere utilizzati tre metodi: clonaggio standard, clonaggio Golden Gate e assemblaggio in fase solida o FLASH.

In generale, per clonaggio si intendono un insieme di metodi che consentono di generare molecole ricombinanti e ottenere numerose copie di esse ossia cloni. Per fare ciò, il frammento di interesse viene isolato con enzimi di restrizione e poi inserito in un vettore di espressione grazie a reazioni di ligazione. La molecola ricombinante finale, cioè il frammento di interesse contenuto nel vettore scelto, può essere trasfettata in cellule ospiti, nelle quali verrà espresso, oppure può essere sottoposta a PCR. In entrambe le modalità si ottengono molte copie della molecola di interesse, la quale può essere di natura nucleotidica o proteica (geni, mRNA, proteine).

Il *clonaggio standard* si basa sull'utilizzo di plasmidi, contenenti i frammenti di DNA codificanti le singole ripetizioni TALE, e sull'unione di essi attraverso reazioni di digestione enzimatica e di ligazione. Attualmente in commercio si trovano quattro tipologie di moduli (ripetizioni), ognuno contenuto in plasmidi differenti, con residui iper-variabili NN, NI, HD e NG che riconoscono rispettivamente guanina, adenina, citosina e timina. In tutti questi plasmidi i frammenti di DNA codificanti i moduli sono fiancheggiati, al 3' e al 5', da due siti di restrizione per endonucleasi differenti. In particolare, questi enzimi sono in grado di riconoscere siti bersaglio differenti ma di tagliare generando estremità compatibili. La digestione dei frammenti, eseguita dagli enzimi di restrizione scelti, è seguita dalla ligazione, processo che porta all'unione delle ripetizioni. La strategia di ligazione segue uno schema per ottenere un dominio TALE formato dai quattro moduli ordinati in modo tale da riconoscere la sequenza target. Tale strategia è basata sulla ligazione dell'estremità 3' di un frammento di DNA, codificante un modulo, con l'estremità 5' di un secondo frammento. Il risultato sarà un frammento di DNA, codificante i due moduli, caratterizzato da estremità 3' e 5' compatibili con ulteriori moduli. In questo modo sarà soggetto a ligazioni successive per generare un frammento di DNA che codificherà il dominio TALE finale. Il vantaggio principale di questo approccio è che il clonaggio è un metodo utilizzato nella maggior parte dei laboratori ma è una procedura laboriosa che richiede diverse settimane; quindi, non è applicabile per creare grandi quantità di domini TALE.

Il *clonaggio Golden Gate* permette di unire in una singola reazione fino a 10 frammenti di DNA, ognuno codificante una ripetizione di TALE, in un determinato ordine all'interno di un vettore plasmidico. Per fare ciò, i frammenti da legare vengono sottoposti a digestione enzimatica per ottenere estremità sporgenti al

5'. Grazie alle diverse estremità sporgenti create è possibile unire frammenti diversi in un ordine ben preciso. Il principale limite di questo approccio è che possono essere uniti contemporaneamente e in modo efficiente un numero massimo di 10 frammenti costruendo dei plasmidi che codificano sottoinsiemi di domini ripetuti. Tali intermedi poi verranno ulteriormente uniti tra loro per ottenere il dominio TALE finale. Il metodo Golden Gate, rispetto al clonaggio standard, permette di ridurre i tempi di realizzazione a circa una settimana.

Entrambi i metodi sopra descritti non permettono una produzione automatica ad alto rendimento e consentono la produzione di array con un limite massimo di ripetizioni. Da una decina di anni è emerso il metodo di *assemblaggio FLASH* (Fast ligation-based Automatable Solid-phase High-throughput) che permette la costruzione di domini TALE su larga scala. Tale metodica si basa sulla creazione di 376 plasmidi che codificano uno, due, tre o quattro ripetizioni TALE costituiti da tutte le combinazioni possibili dei quattro domini presenti in commercio (descritti precedentemente). Secondo questo procedimento, il primo frammento di DNA del dominio TALE finale, codificante la/le ripetizione/i, presenta all'estremità 5' una molecola di biotina. Quest'ultima è caratterizzata da un'elevata affinità per la proteina streptavidina favorendo, così, il legame del primo frammento ad una sfera magnetica rivestita di tale molecola proteica. L'estremità 3' del primo frammento verrà sottoposta a digestione enzimatica e a successiva ligazione con altri frammenti, codificanti le ripetizioni pre-assemblate, fino a quando non viene generato l'array desiderato. Una volta ottenuto il frammento di DNA codificante l'intero dominio TALE, esso verrà sottoposto a digestione enzimatica all'estremità 5' in modo tale da scinderlo dalla sfera magnetica.

L'intero processo può essere svolto da uno strumento robotizzato all'interno di una piastra da 96 pozzetti oppure manualmente con l'uso di una pipetta multicanale. Il vantaggio del metodo FLASH è che i frammenti di DNA vengono assemblati su sfere magnetiche, dunque in fase solida, permettendo l'esecuzione di reazioni enzimatiche seriali senza la necessità di fare lavaggi o purificazioni, evitando la creazione di costrutti intermedi. Inoltre, permette di realizzare un array di qualsiasi lunghezza desiderata.

Una volta ottenuto il frammento di DNA, codificante il dominio TALE, sono state ottenute le ORF di DdCBE come blocchi genici. Questi ultimi sono stati realizzati da GeneArt, azienda biotecnologica tedesca che sintetizza costrutti di DNA personalizzabili. Tali ORF poi sono state inserite all'interno del plasmide pVAX a valle della sequenza di localizzazione mitocondriale, MTS. Quest'ultima deriva dalla superossido dismutasi mitocondriale (SOD2), ossia un enzima in grado di convertire l'anione superossido, prodotto nella catena respiratoria, in perossido di idrogeno. La sequenza MTS è stata ottenuta per digestione enzimatica dal frammento di DNA codificante SOD2. Il plasmide finale è stato poi amplificato attraverso PCR e utilizzato per costruire le particelle virali di AAV.

4.2 Estrazione genoma mitocondriale

Il mtDNA è stato ottenuto dai mitocondri, isolati dai campioni di tessuto cardiaco, utilizzando kit di estrazione derivanti da Promega.

4.3 High-throughput sequencing

Il sequenziamento ad alta resa, conosciuto anche come Next Generation Sequencing (NGS), rappresenta un insieme di metodiche di sequenziamento che consentono di analizzare un'elevata quantità di sequenze in parallelo in breve tempo. Le due caratteristiche principali di questi processi sono (1) il sequenziamento avviene secondo la tecnica "sequencing by synthesis" in quanto le molecole vengono contemporaneamente copiate e sequenziate (2) la realizzazione, per ogni molecola da analizzare, di cloni in modo tale da generare un segnale di fluorescenza tale da essere rilevato dal detector. Sotto il nome di NGS, sono raggruppati diverse tipologie di sequenziatori suddivisi in categorie denominate "prima generazione", "seconda generazione" (Illumina) e "terza generazione" (Nanopore e PacBio). Questa tesi tratterà del sequenziamento Illumina in quanto è stato utilizzato dagli autori dell'articolo per sequenziare la regione target dell'editor DdCBE in seguito al suo utilizzo. Lo scopo è quello di rilevare se le citosine d'interesse sono state convertite in uracile e anche se altre (citosine off-target) sono state modificate.

Illumina si basa su cinque passaggi fondamentali: preparazione del DNA templato, amplificazione delle singole molecole attraverso la tecnica bridge PCR, sequenziamento, acquisizione delle immagini di fluorescenza e ricostruzione della sequenza. Per la preparazione del DNA templato, gli autori hanno amplificato una regione di circa 16bp del gene *MT-ND3* utilizzando due primer forward (5'-3'). Tali primer hanno una lunghezza di 30 bp e sono stati progettati per essere complementari ad una regione a monte e a valle della sequenza templato localizzata nel filamento 3'-5'. Al termine della reazione di PCR, l'amplicone ottenuto è stato poi sottoposto a ligazione con adattatori in entrambe le estremità. A questo punto, le molecole di amplicone ottenute saranno il templato del bridge PCR. Infatti, esse vengono depositate sulla superficie della flow cell alla quale si trovano legati covalentemente i primer del bridge PCR che, però, sono complementari agli adattatori legati al DNA templato. Questo comporta il legame di entrambe le estremità del templato al supporto solido facendo assumere alla molecola una forma a ponte, dalla quale prende il nome questa PCR. Una volta depositato il DNA templato e aggiunti i reagenti, il bridge PCR segue lo stesso processo della PCR classica. Dunque, il primer viene esteso dalla polimerasi copiando il filamento stampo originando, così, una molecola a doppio filamento. In seguito, avviene la denaturazione che porta all'apertura della struttura a ponte. La ripetizione delle fasi di sintesi e di denaturazione porta alla generazione di un cluster di molecole uguali al DNA templato iniziale. Il cluster, una volta reso a singolo filamento, sarà sottoposto alla fase di sequenziamento che avviene con la tecnica "sequencing by synthesis".

Tale strategia si basa sull'estensione di un oligonucleotide, che funge da innesco, copiando il filamento stampo con l'incorporazione, da parte della polimerasi, di nucleotidi associati ad un gruppo fluorescente diverso in funzione della base azotata. I nucleotidi, oltre a possedere un gruppo fluorescente, presentano all'estremità 3' un gruppo blocco che impedisce l'aggiunta momentanea di un secondo nucleotide. Questo perché la procedura prevede, dopo ogni aggiunta, la rilevazione della fluorescenza emessa dal nucleotide che è stato incorporato da un detector. Tale informazione poi viene trasmessa ad un software, il quale, registrando le diverse fluorescenze rilevate ad ogni ciclo di sequenziamento, le converte nella sequenza nucleotidica corrispondente.

4.4 Real-time PCR

La real-time PCR è una tecnica che permette di determinare la quantità di templatato iniziale attraverso il calcolo del ciclo soglia. La procedura sperimentale è pressoché identica a quella della PCR classica; infatti, la reazione prevede la ripetizione di 30-35 cicli ognuno dei quali è costituito da tre fasi: denaturazione, annealing ed estensione. Brevemente, la prima fase prevede il riscaldamento del campione ad una temperatura di circa 96° in modo tale da denaturare la doppia molecola, nella seconda il templatato viene portato alla temperatura di annealing (circa 55°), temperatura alla quale i primer si appaiano alla sequenza di DNA complementare del templatato mentre la terza e ultima fase prevede l'estensione del primer copiando il filamento stampo da parte della polimerasi. La real-time PCR differisce dalla PCR classica in quanto la rilevazione della quantità di amplicone avviene al termine di ogni ciclo, ottenendo così la relazione tra la quantità di fluorescenza prodotta e il numero di cicli. Questa relazione ha un andamento sigmoidale e, quindi, è caratterizzata da una fase dove la fluorescenza aumenta in maniera esponenziale. Solo all'interno di questa fase, la quantità di amplicone prodotto è proporzionale alla quantità di templatato iniziale. È per questo motivo che un software, a partire da questo andamento, estrae il ciclo soglia (Ct), definito come il numero di cicli necessari per generare un segnale di fluorescenza rilevabile. Il ciclo soglia da informazioni sulla quantità di templatato perché dipende in maniera proporzionale da esso. Infatti, maggiore è la quantità di templatato iniziale e minore sarà il ciclo di PCR al quale la fluorescenza diventa rilevabile.

Gli autori dell'articolo hanno utilizzato la real-time PCR per quantificare le copie dei genomi virali (AAV-DdCBE-L e AAV-DdCBE-H) e mtDNA nei campioni di cuore. Per entrambi i casi, sono stati impiegati due primer (forward e reverse) e una sonda. Quest'ultima è un oligonucleotide progettato per essere complementare ad una porzione dell'amplicone, diminuendo così la rilevazione di eventuali aspecifici.

5. DISCUSSIONE E CONCLUSIONE

Gli autori dell'articolo su cui si basa questa tesi, si sono posti l'obiettivo di progettare uno strumento che consenta di modificare il mtDNA. Si tratta di DdCBE, un editor di citosina derivante da una DNA-deaminasi batterica (DddA), in grado di modificare tale base azotata in uracile mediante deaminazione. Per valutare l'efficienza di DdCBE, l'esperimento è stato condotto in vivo in topi adulti e neonati trasmettendo l'editor nel tessuto cardiaco mediante AAV. In entrambi i casi, sono stati ottenuti tre campioni di tessuto cardiaco derivanti da (1) topi di controllo nei quali è stato iniettato solo il vettore (2) topi nei quali è stata iniettata la coppia 1 di DdCBE (3) topi nei quali è stata somministrata DdCBE cataliticamente inattiva. Nel caso dell'esperimento riguardante i topi adulti, una parte di essi sono stati sacrificati dopo 3 settimane mentre i restanti dopo 24 settimane dalla somministrazione di DdCBE. I topi neonati sono stati sacrificati tutti dopo 3 settimane dalla somministrazione. Ritengo che sia corretto mantenere costante, tra adulti e neonati, l'intervallo di tempo tra la somministrazione e la soppressione in modo tale che la sola variabile dell'esperimento sia l'età del topo. Con questo intendo dire che, se questo intervallo fosse diverso tra adulti e neonati, per il gruppo che ha un intervallo maggiore probabilmente verrebbero rilevate una quantità di citosine modificate maggiore rispetto a quelle che verrebbero modificate se il tempo fosse uguale. Nonostante ciò, avrei sacrificato un gruppo di topi neonati anche dopo 24 settimane in modo tale da ottenere dei risultati, rappresentativi i neonati dopo 24 settimane, da confrontare con quelli degli adulti.

I risultati ottenuti dagli autori dello studio dimostrano che DdCBE è in grado di modificare le citosine sia nei topi adulti che neonati. In particolare, considerando gli adulti, il numero di citosine C12 e C13 deaminate in uracile dopo 24 settimane (10-20%) è maggiore rispetto a quelle rilevate dopo 3 settimane (1-2%). Questo perché l'editor ha avuto a disposizione un tempo maggiore, deaminando così un maggior numero di citosine. Osservando i risultati ottenuti dai topi neonati (20-30%) dopo 3 settimane, si nota che la quantità di citosine d'interesse convertite in uracile è maggiore di quella degli adulti sia dopo 3 settimane sia dopo 24 settimane. Ci sono due motivazioni che spiegano ciò (1) la dimensione ridotta del cuore dei neonati favorisce l'infezione. Infatti, l'elevato rapporto tra il vettore AAV, contenente DdCBE, e le cellule cardiache agevola la trasduzione di DdCBE nel tessuto cardiaco (2) lo sviluppo postnatale del cuore prevede l'aumento della dimensione delle cellule cardiache grazie alla sintesi proteica, dunque, il mtDNA va incontro a numerose replicazioni. È stato stimato che, durante le prime 3 settimane di vita, il numero di copie di mtDNA aumenta di circa 13 volte. Questa finestra temporale, nella quale il mtDNA subisce molte replicazioni, determina l'aumento della probabilità di fissare gli eventi di deaminazione della citosina in timina.

Ci sono due aspetti che potrebbero impedire un efficiente editing da parte di DdCBE (1) il sistema DdCBE comporta l'inibizione di UGI. Questo porta alla rilevazione di tutte le citosine deaminate in uracile nella sequenza target anche se non derivano dall'azione dell'editor (2) l'intervallo di tempo tra la somministrazione e la soppressione dei topi deve essere sufficiente per garantire una quantità di replicazioni tali da garantire la fissazione della deaminazione della citosina in timina. Questo studio dimostra che 3 e 24 settimane sono sufficienti per tale scopo. Questi risultati sono molto rilevanti dal punto di vista della terapia delle malattie mitocondriali dovute a mutazioni del mtDNA perché, riducendo il carico mutazionale al di sotto della soglia critica, si può ottenere un effetto terapeutico anche se permane una quota di mtDNA mutato.

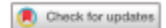
Un vero limite di DdCBE è la presenza di editing al di fuori della sequenza bersaglio, sia nei topi adulti sia neonati. Gli autori dell'articolo hanno testato anche questo aspetto e hanno osservato un editing off-target che, negli adulti dopo 24 settimane dalla somministrazione, risulta 7 volte superiore rispetto ai controlli (somministrati DdCBE cataliticamente inattivi) mentre nei neonati è 3 volte superiore a quello ottenuto dagli adulti. Questo dimostra la presenza di una correlazione positiva tra on-target e off-target.

La presenza di un'attività off-target suggerisce che lo studio di DdCBE in futuro deve concentrarsi sul miglioramento della precisione di questo strumento in modo tale da ridurre il più possibile questo effetto indesiderato. In seguito, DdCBE potrebbe essere utilizzato nella terapia genica per la correzione delle mutazioni responsabili di malattie mitocondriali. In particolare, questo studio ha dimostrato che l'efficienza di DdCBE è maggiore nei topi neonati rispetto agli adulti, quindi, potrebbe essere impiegato per curare le malattie ad insorgenza postnatale.

6. BIBLIOGRAFIA

1. Gorman, Gráinne S., et al. "Mitochondrial Diseases." *Nature Reviews Disease Primers*, vol. 2, no. 1, 20 Oct. 2016.
2. Joung, J. Keith, and Jeffrey D. Sander., et al. "TALENs: A Widely Applicable Technology for Targeted Genome Editing." *Nature Reviews Molecular Cell Biology*, vol. 14, no. 1, 21 Nov. 2012, pp. 49–55.
3. Mok, Beverly Y., et al. "A Bacterial Cytidine Deaminase Toxin Enables CRISPR-Free Mitochondrial Base Editing." *Nature*, vol. 583, no. 7817, July 2020, pp. 631–637.
4. Reyon, Deepak, et al. "FLASH Assembly of TALENs for High-Throughput Genome Editing." *Nature Biotechnology*, vol. 30, no. 5, 8 Apr. 2012, pp. 460–465.
5. Silva-Pinheiro, Pedro, et al. "In Vivo Mitochondrial Base Editing via Adeno-Associated Viral Delivery to Mouse Post-Mitotic Tissue." *Nature Communications*, vol. 13, no. 1, 8 Feb. 2022, p. 750.
6. Stewart, James B., and Patrick F. Chinnery, et al. "The Dynamics of Mitochondrial DNA Heteroplasmy: Implications for Human Health and Disease." *Nature Reviews Genetics*, vol. 16, no. 9, 18 Aug. 2015, pp. 530–542.



ARTICLE



<https://doi.org/10.1038/s41467-022-28358-w>

OPEN

In vivo mitochondrial base editing via adeno-associated viral delivery to mouse post-mitotic tissue

Pedro Silva-Pinheiro ¹, Pavel A. Nash¹, Lindsey Van Haute ¹, Christian D. Mutti ¹, Keira Turner ¹ & Michal Minczuk ¹✉

Mitochondria host key metabolic processes vital for cellular energy provision and are central to cell fate decisions. They are subjected to unique genetic control by both nuclear DNA and their own multi-copy genome - mitochondrial DNA (mtDNA). Mutations in mtDNA often lead to clinically heterogeneous, maternally inherited diseases that display different organ-specific presentation at any stage of life. For a long time, genetic manipulation of mammalian mtDNA has posed a major challenge, impeding our ability to understand the basic mitochondrial biology and mechanisms underpinning mitochondrial disease. However, an important new tool for mtDNA mutagenesis has emerged recently, namely double-stranded DNA deaminase (DddA)-derived cytosine base editor (DdCBE). Here, we test this emerging tool for in vivo use, by delivering DdCBEs into mouse heart using adeno-associated virus (AAV) vectors and show that it can install desired mtDNA edits in adult and neonatal mice. This work provides proof-of-concept for use of DdCBEs to mutagenize mtDNA in vivo in post-mitotic tissues and provides crucial insights into potential translation to human somatic gene correction therapies to treat primary mitochondrial disease phenotypes.

¹MRC Mitochondrial Biology Unit, University of Cambridge, Cambridge CB2 0XY, UK. ✉email: michal.minczuk@mrc-mbu.cam.ac.uk

Mitochondria play a central role in energy provision to the cell and in several key metabolic pathways, such as thermogenesis, calcium handling, iron-sulfur cluster biogenesis, and apoptosis^{1,2}. Mitochondria produce energy in the form of ATP, which is synthesized in the process of oxidative phosphorylation (OXPHOS), involving sequential redox reactions coupled with proton pumping performed by mitochondrial membrane-embedded respiratory chain complexes (I–IV) and ATP synthase (complex V). In mammals, the mitochondrial proteome comprises ~1200 proteins³, with most of them being nuclear DNA (nDNA)-encoded. However, 13 essential OXPHOS polypeptides, and 22 tRNAs, and 2 rRNAs required for their translation reside inside the mitochondrial matrix, encoded by the mitochondrial DNA (mtDNA)—a 16.5 kb, maternally inherited multicopy circular genome.

Mitochondrial diseases are genetic disorders, caused by mutations, either in nDNA or mtDNA, that lead to mitochondrial energy production impairment and perturbations in other aspects of cellular homeostasis. With a prevalence of ~23 in 100,000, mitochondrial disorders are among the most common inherited diseases, and are often associated with severe disability and shortened lifespan⁴. There are currently no effective treatments for these disorders and clinical management focuses on treating complications⁵. Mutations in mtDNA and mitochondrial dysfunction have also been implicated in many common diseases with high societal impact and ageing^{6,7}. Mammalian cells can contain 100–1000 copies of mtDNA⁸. Pathogenic variants in mtDNA can either be present in all copies (homoplasmy) or only in a portion of genomes (heteroplasmy), with mutant load varying across cells, tissues, and organs⁹. In heteroplasmic cells, the mutant load required for clinical expression must exceed a threshold, which is highly variable and is dependent on the mtDNA variant, affected tissues/organ(s), and genetic/environmental contexts, but is usually more than 60%¹⁰.

Despite the ongoing genome-engineering revolution enabled by the CRISPR/Cas-based systems, mammalian mtDNA has been resistant to genetic modifications, in a large part owing to ineffective nucleic acid import into the mitochondria¹¹. The inability to edit mtDNA sequences in mammalian mitochondria within cells has hampered the research of normal mtDNA processes, the development of in vivo models and therapies for mtDNA diseases. For many years the approaches towards manipulation of mtDNA in mammals have been mainly limited to mitochondrially targeted restriction enzymes^{12–14} and programmable nucleases^{15–21}. These nucleases have been used to eliminate undesired mtDNA molecules from heteroplasmic populations, to move the mutant mtDNA heteroplasmy below the pathogenicity threshold²². Following extensive trials in vitro, the mitochondrial nuclease-based approaches have reached in vivo proof-of-concept. Delivered by adeno-associated virus (AAV) in heteroplasmic mouse models, mtRE, mtZFN, mitoTALEN, or mitoARCUS demonstrated specific elimination of the mutant mtDNA in the target tissues^{23,24}, which in some models and approaches was accompanied by the molecular and physiological rescue of disease phenotypes^{25–27}.

While programmable nucleases have proven useful in changing the existing heteroplasmy, they are unable to introduce novel mtDNA variants. However, recently a novel tool has emerged: DddA-derived cytosine base editor (DdCBE), which catalyzes site-specific C:G to T:A conversions in mtDNA with good target specificity in human cultured cells²⁸. DdCBE is based upon a modified bacterial toxin DddA_{tox} (separated, non-toxic halves fused to TALE proteins) which is targeted to the mitochondrial matrix to catalyze the deamination of cytidines within dsDNA at sequence determined by TALE design²⁸. Current DdCBEs deaminate cytidines (in the TC:GA sequence context) to uracil

leading to a TC:GA > TT:AA mutations upon subsequent replication²⁸. An initial proof-of-concept of successful installment of mtDNA edits by delivering DdCBEs mRNA into embryos in the mouse²⁹ and zebrafish³⁰ has also recently been provided.

In this study, we provide proof-of-concept for the use of DdCBEs in vivo in somatic tissue. We used the mouse heart as a surrogate of a post-mitotic tissue and showed that DdCBE delivered using AAV can install the desired mtDNA mutations in adult and neonatal mice. To the best of our knowledge, such a result has not been reported in the literature thus far. This work demonstrates that the DdCBE platform could be used for tissue-specific mtDNA mutagenesis in vivo and potentially for future therapies based upon somatic mitochondrial gene correction to treat mtDNA-linked mitochondrial diseases.

Results

Design of DdCBE and mtDNA editing in mouse cultured cells.

With the intention of testing the emerging mtDNA editing DdCBE technology, we set off to induce de novo mutations in mouse mitochondrial complex I in cultured cells and in somatic tissues upon AAV delivery. We aimed at editing the GGA glycine 40 codon in mouse *MT-Nd3* (mtDNA positions: m.9576 G and m.9577 G) by targeting the complementary cytosine residues with DdCBE (Fig. 1a, C₁₂ C₁₃). To enable this, we designed four DdCBE pairs, containing TALE domains binding the mtDNA light (L) and heavy (H) strands (mtDNA positions m.9549–m.9564 and m.9584–m.9599, respectively) and different combinations of DddA_{tox} splits (G1333 or G1397), targeting a 19 bp-long sequence in the mouse *MT-Nd3* gene (mtDNA positions: m.9565–m.9583) (Fig. 1a–b) and named them DdCBE-Nd3-9577-1 to 4. The previous study has shown that for the sites containing two consecutive cytosines (preceded by a thymine), both of these cytosines can be edited by DdCBEs, with the following potential consequences TCC:GGA > TTC:GAA, TCC:GGA > TCT:AGA, TCC:GGA > TTT:AAA (edited C underlined)²⁸. In line with this, we predicted three possible outcomes of m.9576 G and m.9577 G editing (Fig. 1b): [i] deamination of both complementary cytosines (Fig. 1b, C₁₂ and C₁₃), would result in glycine to lysine mutation (G40K), [ii] deamination of C₁₃ would lead to a glycine to glutamic acid mutation (G40E), whereas [iii] exclusive editing of C₁₂ would lead to a premature AGA stop codon (G40*), according to the mitochondrial genetic code (Fig. 1b). The three predicted mutations are located in the conserved ND3 loop, involved in active/deactive state transition of complex I (Fig. 1c–d)^{31,32}.

Next, we transiently delivered DdCBE-Nd3-9577 pairs into mouse cultured NIH/3T3 cells and selected the transfectants using fluorescence-activated cell sorting (FACS) at 24 h post-transfection and allowed for a 6 day-long recovery (Fig. 2a). After 7 days, we detected efficient editing of target cytosines by Sanger sequencing for the DdCBE-Nd3-9577 pairs 1, 2 and 3 (Fig. 2b). Although the targeted 19 bp sequence region between the TALE binding sites contained five cytosine residues with the correct thymine-cytosine (TC) consensus (Fig. 1a, C₆, C₁₂, C₁₃, C₁₇, and C₁₈), positioning of the DddA_{tox} deaminase domain allowed for editing of C₁₂, C₁₃ leading to the expected changes in the *MT-Nd3* GGA glycine 40 codon (Fig. 2b). Next, we performed next-generation sequencing (NGS) analysis of the DdCBE-Nd3-9577-mediated mutagenesis to quantify the editing in each target cytosine and measure the proportion of editing resulting in the G40K, G40E, G40* mutations. This analysis revealed up to ~43% editing of C₁₂ and C₁₃ for DdCBE-Nd3-9577 pair 1, ~20–35% editing for pairs 2 and 3, and confirmed negligible editing activity of pair 4 (Fig. 2c). Furthermore, most of the NGS reads for pair 1 contained editing of both C₁₂ and C₁₃ translating to G40K mutation (~92.5%), whereas DNA reads corresponding to the

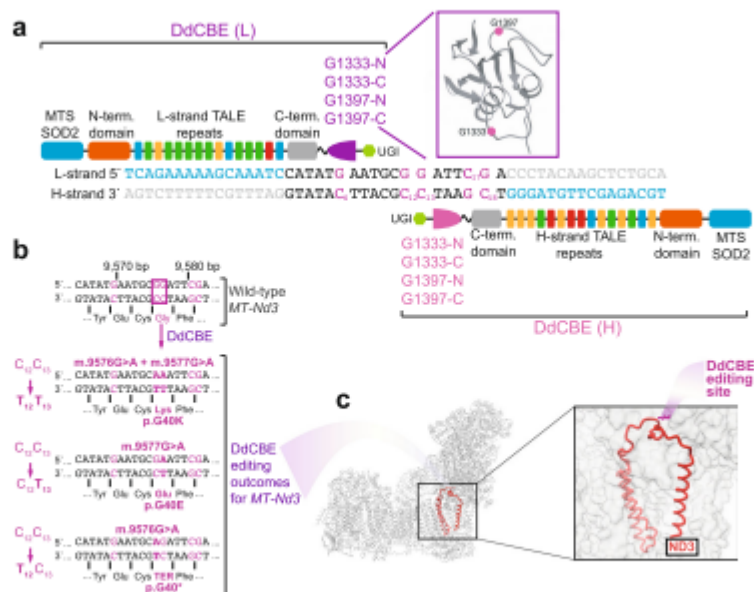


Fig. 1 Design of DdCBEs and mutagenesis site. **a** The architecture of DdCBE monomers targeting m.9576 G (C₁₂) and m.9577 G (C₁₃). The DNA specificity is provided by TALE domains. In each experiment, different DddA_{tox} splits are used (G1397 or G1333, purple) to achieve editing of “TC” sites. MTS SOD2 mitochondrial targeting sequence from superoxide dismutase 2, UGI uracil glycosylase inhibitor, L-strand or (L) light mtDNA strand, H-strand or (H) heavy mtDNA strand. **b** The details and possible outcomes of m.9576 G (C₁₂) and m.9577 G (C₁₃) editing. The purple box indicates the desired editing sites; other potential editing sites are indicated in purple. **c** The structural model of mouse complex I with indicated MT-ND3 subunit (red). The inset shows the location of MT-ND3 p.G40K mutation on the evolutionary conserved MT-ND3 loop.

G40E and G40* changes constituted only ~4.5% and 3%, respectively (Fig. 2d). A similar editing pattern, but with a higher proportion of reads corresponding to G40E, was observed for pair 3 (G40K: ~83%, G40E: 14% G40*: 3%) (Fig. 2d). However, the mutagenesis pattern detected for pair 2 was skewed towards G40E, as compared to pairs 1 and 3, with reads corresponding to G40E accounting for 45.5% of C₁₂ and C₁₃-edited reads (G40K: ~53% and G40*: 1.5%) (Fig. 2d). The NGS analysis revealed a low level (below 3%) of C₆ (mtDNA position m.9570 G) for pairs 1–3, which is predicted to install the E38K change (Fig. 2c). To confirm that the observed mtDNA editing is indeed a result of the catalytic activity of DddA_{tox}, we used the DdCBE-Nd3-9577 pairs harboring a catalytically inactive DddA_{tox} (E1347A) in a control transient transfection experiment. The NGS analysis showed that none of the catalytically inactive DdCBEs exerted detectable deamination activity that would lead to mtDNA editing, with the mutagenesis frequency being at the level of wild-type cells (Fig. 2c). Based on these in vitro results, we concluded that the DdCBE-Nd3-9577-1 pair is the most suitable for efficient installation of G40K and decided to proceed with in vivo experiments using this set.

AAV-based in vivo DdCBE editing of mtDNA in adult mice. To provide a proof-of-concept for in vivo mtDNA gene editing of somatic cells, we encapsidated the catalytically active and inactive versions of the DdCBE-Nd3-9577-1 monomers into the cardiotropic AAV9.45 serotype and administered them systemically via tail-vein injection at 1×10^{12} viral genomes (vg) per monomer per 8-week-old adult mouse (Fig. 3a). At 3- and 24-weeks post-injection, we confirmed successful DdCBE DNA delivery to the cardiac tissue by quantitative PCR (Supplementary Fig. 1a–b) and detected

its expression in total mouse heart tissue by western blotting (Supplementary Fig. 1c) and immunohistochemistry (Supplementary Fig. 1d–e). At 3 weeks after DdCBE-Nd3-9577-1 AAV injections, we detected low-level editing (1–2%) of the target C₁₂ and C₁₃ bases (corresponding to m.9576G > A and m.9577G > A) by NGS, but not by Sanger sequencing (Fig. 3b–c). The NGS analysis revealed C₁₂ and C₁₃ mutagenesis in the hearts of animals injected with catalytically active DdCBE-Nd3-9577-1, but not in those injected with a vehicle or catalytically inactive DdCBE, with the C₁₂ and C₁₃ editing pattern resembling the one observed in in vitro experiments (Fig. 3d). Notwithstanding, at 24-weeks post-injection robust editing (10–20%) of C₁₂ and C₁₃ was observed in the cardiac tissue of mice injected with catalytically active DdCBE-Nd3-9577-1 by Sanger sequencing and NGS, with no detectable changes being detected for vehicle and catalytically inactive controls (Fig. 3e–f). Further analysis confirmed that in the majority (94%) of the NGS reads both C₁₂ and C₁₃ were edited, translating into the G40K MT-ND3 mutation (Fig. 3g). We did not detect any adverse effect on mtDNA copy number in DdCBE AAV-transduced cardiac tissue as compared to vehicle-injected controls (Supplementary Fig. 1f–g). Taken together, these results established that DdCBE editing can be applied to install mtDNA mutations in vivo in post-mitotic tissues.

AAV-based in vivo DdCBE editing of mtDNA in neonates. Having established that de novo mutations can be installed in vivo in the post-mitotic tissue of adult mice following AAV-assisted DdCBE transduction, we set off to investigate whether transgene delivery to younger subjects could enhance the editing. To this end, we injected neonatal subjects (first 24 h of life) with DdCBE-Nd3-9577-1 AAV9.45 and its catalytically inactive version via temporal vein injection at 1×10^{12} vg per monomer per animal (Fig. 4a).

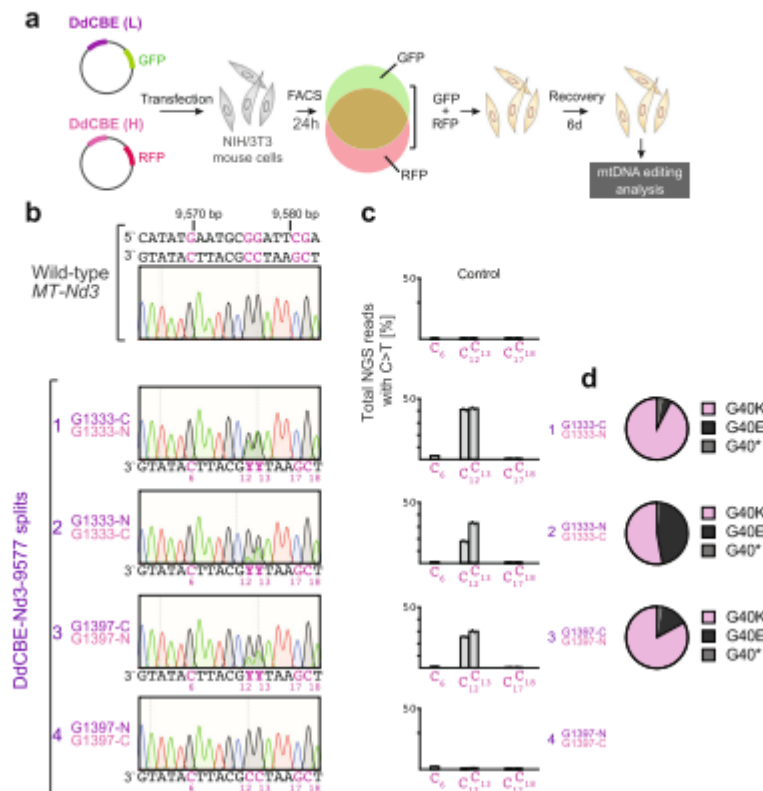


Fig. 2 Mitochondrial DNA editing in cultured mouse cells. **a** Schematic of the general workflow for in vitro experiments that involve transient transfection of cultured mouse NIH/3T3 cells with plasmids co-expressing DdCBE monomers and fluorescent marker proteins, FACS-based selection of cells expressing both monomers and evaluation of mtDNA from DdCBE-treated cells. **b** Editing of mouse *MT-Nd3* with different DdCBE splits in cells 7 days post-transfection, analyzed by Sanger sequencing. Potential editing sites are indicated in purple. The C:G > T:A deamination leads to m.9576 G > A ($C_{12} > T_{12}$) and m.9577 G > A ($C_{13} > T_{13}$) mutations, which translate to p.G40K change in *MT-ND3*. **c** The NGS analysis of the editing region in cells treated with different DdCBE splits. Bars represent the mean ($n = 2$). Source data are provided as a Source Data file. The mutagenesis frequency for the catalytically inactive versions is provided in the Source Data file. **d** The distribution of NGS reads containing m.9576 G (C_{12}) or m.9577 G (C_{13}) edits. The G40K reads contain both m.9576 G > A ($C_{12} > T_{12}$) and m.9577 G > A ($C_{13} > T_{13}$) mutations, G40E reads contain only the m.9577 G > A ($C_{13} > T_{13}$) mutation, while G40* reads contain only the m.9576 G > A ($C_{12} > T_{12}$) mutation. Source data are provided as a Source Data file.

Upon sacrificing the animals at 3-weeks post-injection, we confirmed successful DdCBE-Nd3-9577-1 AAV delivery (Supplementary Fig. 2a) and robust expression in the total mouse heart (Supplementary Fig. 2b-c). At this time-point we observed high efficiency of editing of mtDNA in mouse heart, with Sanger sequencing and NGS revealing 20–30% of C > T (G > A) changes of C_{12} and C_{13} in the DdCBE-Nd3-9577-1-targeted spacing region (Fig. 4b-c). No editing of these bases was observed in the control pups injected with a vehicle and the catalytically inactive DdCBE-Nd3-9577-1 AAVs (Fig. 4c). The vast majority of NGS reads (95%) contained simultaneous edits of C_{12} and C_{13} (corresponding to m.9576 G > A and m.9577 G > A), translating into the G40K *MT-ND3* mutation (Fig. 4d). We did not observe any significant changes in mtDNA copy number in the AAV-treated mice as compared to the vehicle-injected controls (Supplementary Fig. 2d). Taken together, these data not only further confirm that DdCBE-mediated mtDNA editing is possible in post-mitotic tissues upon AAV delivery, but also show that treatment of the younger subjects is beneficial for the efficacy of mtDNA modification.

Off-target editing by DdCBE in adult and neonatal mice followed AAV delivery. To score mtDNA-wide off-targeting in the mice, we analyzed mtDNA from hearts of vehicle-injected controls and mice injected with active or inactive versions of the DdCBE-Nd3-9577-1 pair. Vehicle-injected and catalytically inactive editor samples were used as a control in order to distinguish DdCBE-induced C:G-to-T:A single-nucleotide variants (SNVs) from natural background heteroplasmy. The average frequencies of mtDNA-wide off-target C:G-to-T:A editing by DdCBE-Nd3-9577-1 in the adult animals treated for 3 weeks were comparable to those of the vehicle-injected and catalytically inactive-DdCBE controls (0.026–0.046%) (Supplementary Fig. 3a). However, the adult mice treated with DdCBE-Nd3-9577-1 for 24 weeks showed ~7-fold higher average off-target editing frequency (0.22–0.30%) as compared with the controls (Supplementary Fig. 3a). We observed the highest off-target editing in the neonates treated with DdCBE-Nd3-9577-1, which was ~3-fold higher than the one observed for 24-week AAV-treated adult mice (Supplementary Fig. 3a). There was a positive correlation

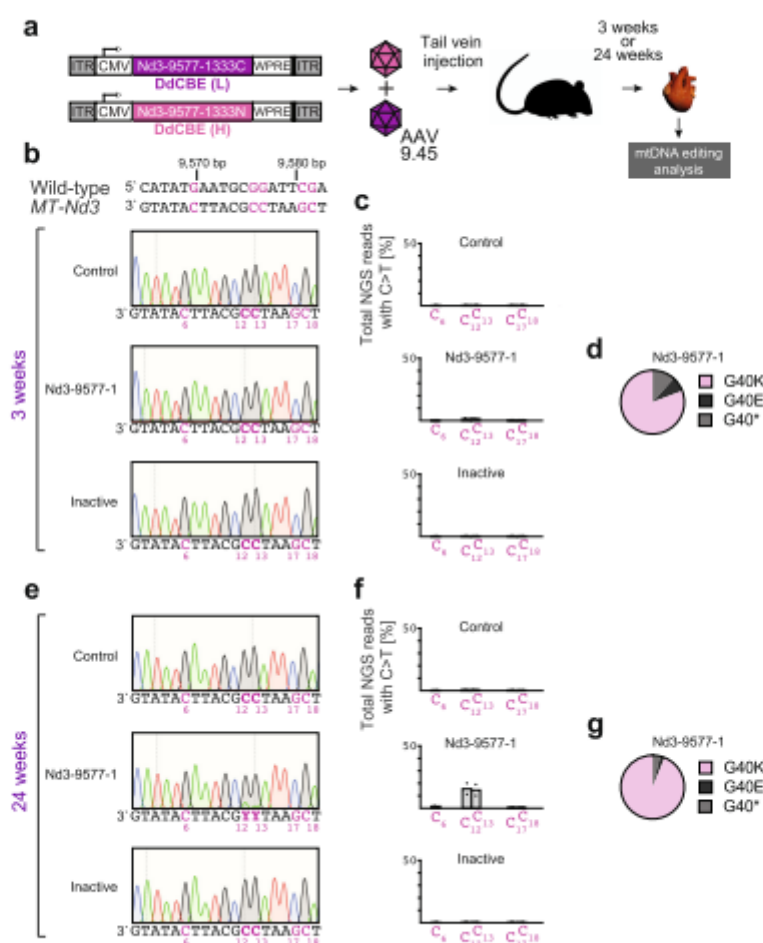


Fig. 3 Mitochondrial DNA editing in adult mouse hearts. **a** Scheme of in vivo experiments with adult mice. The DdCBE-Nd3-9577-1 monomers (see Fig. 2), and their catalytically inactive versions, were encoded in separate AAV genomes, encapsidated in AAV9.45 then simultaneously administered by tail-vein (TV) injection at 1×10^{12} vg/mouse of each monomer. Animals were sacrificed either 3 or 24-weeks post-injection and their cardiac tissue was examined for mtDNA editing. **b, e** Editing of mouse *MT-Nd3* with DdCBE in mouse heart at 3-weeks (**b**) or 24-weeks (**e**) post-injection, analyzed by Sanger sequencing. Potential editing sites are indicated in purple. **c, f** The NGS analysis of the DdCBE editing within the targeted region in adult mouse hearts at 3 weeks (**c**) or 24 weeks (**f**) after injection. Bars represent the mean ($n = 2$). Source data are provided as a Source Data file. **d, g** The distribution of NGS reads containing m.9576 G (C_{12}) or m.9577 G (C_{13}) edits at 3 weeks (**d**) or 24 weeks (**g**) after injection. The G40K reads contain both m.9576 G > A ($C_{12} > T_{12}$) and m.9577 G > A ($C_{13} > T_{13}$) mutations, G40E reads contain only the m.9577 G > A ($C_{13} > T_{13}$) mutation, while G40* reads contain only the m.9576 G > A ($C_{12} > T_{12}$) mutation. Source data are provided as a Source Data file.

between on-target and off-target editing, with increased m.9576 G and m.9577 G modification being accompanied by higher levels of C:G-to-T:A SNVs (Supplementary Fig. 3b). Next, we tested off-target editing at nuclear pseudogenes, which are identical or share a high degree of sequence homology with mtDNA (nuclear mitochondrial DNAs, NUMTs). We did not observe editing beyond the background at the tested NUMTs, even though one of them was identical with the mtDNA on-target sites (Supplementary Fig. 4). This result is consistent with previous reports of exclusive mitochondrial localization of DdCBEs²⁸. Taken together, our data show that, while no detectable off-targeting is observed in nDNA upon AAV delivery of DdCBE, substantial

off-targets are observed in mtDNA, especially when the on-target modification is also high. The latter result suggests that future development of DdCBE must focus on further improvement of the precision of this emerging tool.

Discussion

The discovery of a bacterial cytidine deaminase acting on double-stranded DNA (DddA) led to the development of mitochondrial DddA-derived cytosine base editors (DdCBEs), which is likely to revolutionize the field of mammalian mtDNA genetic modification²⁸. The DdCBE technology provides the potential to reverse engineer the mitochondrial genome in animal cells and

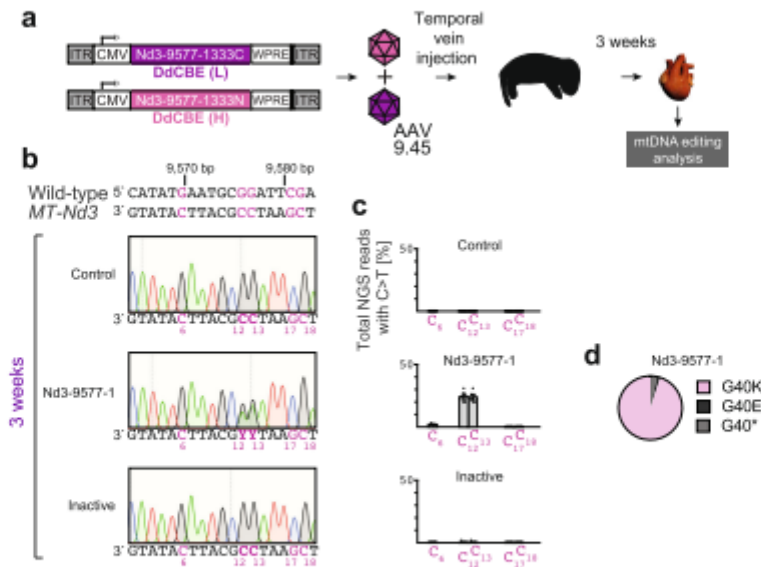


Fig. 4 Mitochondrial DNA editing in neonatal mouse hearts. **a** Scheme of *in vivo* experiments with neonatal mice. The DdCBE-Nd3-9577-1 monomers (see Fig. 2), and their catalytically inactive versions, were encoded in separate AAV genomes, encapsidated in AAV9.45 then simultaneously administered by temporal vein injection at 1×10^{12} vg/mouse of each monomer. Animals were sacrificed 3-weeks post-injection and their cardiac tissue was examined for mtDNA editing. **b** Editing of mouse *MT-Nd3* with DdCBE in neonatal mouse heart at 3-weeks post-injection, analyzed by Sanger sequencing. Potential editing sites are indicated in purple. **c** The NGS analysis of the DdCBE editing within the targeted region in neonatal mouse hearts. Bars represent the mean and error bars represent \pm SEM ($n = 7$). Source data are provided as a Source Data file. **d** The distribution of NGS reads containing m.9576 G (C₁₂) or m.9577 G (C₁₃) edits in neonatal hearts at 3-weeks post-injection. The G40K reads contain both m.9576 G > A (C₁₂ > T₁₂) and m.9577 G > A (C₁₃ > T₁₃) mutations, G40E reads contain only the m.9577 G > A (C₁₃ > T₁₃) mutation, while G40* reads contain only the m.9576 G > A (C₁₂ > T₁₂) mutation. Source data are provided as a Source Data file.

eventually correct homo- and heteroplasmic pathogenic point mutations in mtDNA. Generation of novel animal models is now expected to proceed in an expedited fashion, either by DdCBE-mediated manipulation of Embryonic Stem (ES) cells, direct modification of mouse embryos or, as shown here, by somatic delivery. As with any emerging technologies, the DdCBE approach needs to be tested in multiple systems to validate its usefulness. Thus far, the versatility of DdCBEs was shown by successfully base editing five mtDNA genes with efficiencies ranging between 5 and 50% in human cells *in vitro*²⁸. Further proof-of-concept exemplified the use of base editing in mouse embryos and reported successful germline transmission of DdCBE-induced mtDNA edits²⁹. In the latter report, mutations in *MT-Nd5*, generated by delivering DdCBE mRNAs in mouse zygotes were maintained throughout development and differentiation. These mutations were successfully transmitted to offspring (F1) with heteroplasmy levels of up to 26%, providing evidence that DdCBEs can be used to generate mouse models with bespoke mtDNA mutations²⁹. In this study, we used somatic AAV delivery of DdCBEs, providing a proof-of-concept for an alternative *in vivo* mtDNA mutagenesis means. The presented successful AAV-based mtDNA editing is also critical for *in vivo* proof-of-concept and insights into potential clinical translation to human somatic mitochondrial gene correction therapies to treat primary mitochondrial disease (PMD) phenotypes in worst-affected tissues. These future therapies could be crucial for mtDNA-associated PMDs, whose *de novo* genetics and hard-to-predict penetrance make preimplantation genetic diagnosis (PGD) screening difficult.

We predicted two possible issues that could prevent efficient mtDNA editing by DdCBEs *in vivo*. [i] The editing of mitochondrial genomes with the current DdCBEs involves inhibition of mitochondrial base excision repair (BER), to allow the retention of uracil in DNA (the result of cytosine deamination) and a sufficient level of active mtDNA replication allowing for the conversion of uracil into thymine²⁸. The efficacy of mitochondrial BER had not been fully investigated *in vivo* and it was possible that it operates at the level of preventing efficient DdCBE-mediated editing. [ii] Previous reports highlighted that mammalian mtDNA is replicated continuously even in post-mitotic cells³³, but whether its activity is sufficient to achieve successful editing by DdCBEs was unknown. The successful mtDNA editing in mouse hearts reported here show that [i] the efficacy of BER operating in mouse heart does not lead to immediate excision of uracil in DNA, precluding efficient editing, and [ii] the level of mtDNA replication is high enough for fixing C-to-U deamination events into C-to-T mutations within 3 or 24 weeks after DdCBE treatment of neonatal or adult mice, respectively.

In neonates, our data also show that DdCBE administration earlier during development results in higher editing efficiency. A partial explanation for this finding could be that [i] at earlier vector administration, the greater AAV vector-to-cell ratio promoted higher transduction of the DdCBEs in the neonate heart and [ii] the postnatal development of mouse cardiomyocytes involves substantial mtDNA replication leading to an almost 13-fold increase of mtDNA copy number during the first four weeks of postnatal life³⁴, increasing the probability of fixing C-to-U deamination events into C-to-T mutations. Considering future

therapeutic interventions based on DdCBE-mediated mtDNA correction, this observation may be encouraging in the context of those mitochondrial diseases with early-onset and a rapid progression, for which administration of potential therapies in adults would be inadequate. Nonetheless, further studies on the gene correction potential of AAV-delivered DdCBEs would be required in mtDNA-disease mouse model. However, none of the four currently available mouse models harboring pathogenic mtDNA point mutations is suitable for a DdCBE-mediated gene correction study either due to the nature of mutation (*MT-ND6*: m.13997 G > A, *MT-MK*: m.7731 G > A, *MT-TA*: m.5024 C > T) or incompatible sequence context upstream of a T-to-C mutation (*MT-COI*: m.6589 T > C)³⁵. The latter means that, in addition to DdCBE-based approaches, the field should continue generating novel mouse mtDNA mutated mouse lines using the established, phenotype-first pipeline, as it can generate any mtDNA mutations (not only C:G to T:A offered by DdCBEs)³⁶.

Our results confirm the previous observation that each DddA_{tox} split edits the TC sites with a preference for specific windows in the spacing region. They also highlight that for any given target sequence, testing G1397 and G1333 splits in both orientations are required to achieve on-target editing. Here we demonstrate that the pair DdCBE-Nd3-9577-1 combining the G1333-C split in the TALE targeting the L-strand with G1333-N on the TALE targeting the H-strand produces more robust base editing of C₁₂ and C₁₃ located on the H-strand. Such observations fall in line with the previous reported preference of this split combination, which favors base editing of H-strand Cs present in the center of the editing window between the TALEs²⁸.

Our off-target editing analysis revealed that adult mice treated with DdCBE-Nd3-9577-1 for 24 weeks show ~0.25% of C:G-to-T:A SNV frequencies mtDNA-wide, which is higher than those reported previously for DdCBEs transiently expressed for 3 days in human cells, which ranged between ~0.05 to ~0.15%²⁸. Also, the C:G-to-T:A SNV off-target frequencies detected in the neonates injected with the active DdCBE-Nd3-9577-1 pair for 3 weeks were on average at ~0.8%, which was ~5 times higher than the least “precise” mitochondrial base editor reported by Mok et al.²⁸. We attribute these differences to longer mtDNA-DdCBE exposure time (weeks vs days) in our experiments and conclude that further optimization of mitochondrial DdCBE concentration and specificity will be required, especially in long-term *in vivo* experiments.

In the present study *MT-Nd3* G40K was installed to provide a proof-of-concept of somatic mtDNA editing. However, this mutation is located in the conserved ND3 loop involved in active/deactive state transition^{31,32}. It is expected that high G40K heteroplasmy will result in mitochondrial dysfunction by permanently locking complex I in active confirmation, which warrants further studies. In addition, it has been previously shown that targeting the reversible S-nitrosation of the neighboring ND3 residue C39³⁷ protects against ischaemia-reperfusion (IR) injury. In this line, the G40K mutant could be explored in the context of changes in exposure of Cys39 in models of IR injury.

In conclusion, DdCBE is a promising tool for *de novo* mtDNA editing in post-mitotic tissue, which upon further research and optimization could be used to revert pathogenic mtDNA variants in patients affected with mitochondrial disease.

Methods

Ethics statement. All animal experiments were approved by the local Animal Welfare Ethical Review Body (AWERB) at the University of Cambridge and carried out in accordance with the UK Animals (Scientific Procedures) Act 1986 (Procedure Project Licence: P6C20975A) and EU Directive 2010/63/EU.

Plasmid construction and viral vectors. The DdCBE architectures used were as reported in²⁸. A catalytically dead DddA_{tox} (E1347A) was used in the “inactive” control experiments. TALE arrays were designed using the Repeat Variable

Residues (RVDs) containing NI, NG, NN, and HD, recognizing A, T, G, and C, respectively. To construct the plasmids used in the cell screen, all DdCBEs ORFs were synthesized as gene blocks (GeneArt, Thermo Fisher) and cloned into pVax vectors downstream of a mitochondrial localization signal (MLS) derived from SOD2, using the 5′ KpnI and 3′ BglII restriction sites (Supplementary Sequences 1). Vector construction of DdCBEs intended for AAV production was achieved by PCR amplification of the transgenes to include 5′ NotI and 3′ BamHI sites, allowing cloning into a rAAV2-CMV backbone (Supplementary Sequences 2), previously reported in²⁸. The resulting plasmids were used to generate recombinant AAV2/9.45 viral particles at the UNC Gene Therapy Center, Vector Core Facility (Chapel Hill, NC).

Cell culture and transfections. NIH/3T3 cells (CRL-1658TM, American Type Culture Collection (ATCC)) were cultured at 37 °C under 5 % (vol/vol) CO₂ and in complete Dulbecco’s Modified Eagle Medium (DMEM) (4.5 g/L glucose 2 mM glutamine, 110 mg/ml sodium pyruvate), supplemented with 10% calf bovine serum with iron and 5% penicillin/streptomycin (all from Gibco). Mycoplasma tests in the culture medium were negative. The cell line was not authenticated in this study. For DdCBE pair screen, NIH/3T3 mouse cells plated in six-well tissue culture plates at a confluency of 70% were transfected with 3200 ng of each monomer (L and H), to a total of 6400 ng of plasmid DNA using 16 μl of FuGENE-HD (Promega), following manufacturer’s guidelines. After 24 h, cells were collected for Fluorescence-activated cell sorting (FACS) and sorted for GFP and RFP double-positive cells using a BD FACSMelodyTM Cell sorter. The collected double-positive cells were allowed to recover for another 6 days and then used for DNA extraction, as described below.

Animals. Mice in a C57BL/6J background were obtained from Charles River Laboratories. The animals were maintained in a temperature- and humidity-controlled animal care facility with a 12 h light/12 h dark cycle and free access to water and food, and they were sacrificed by cervical dislocation. In adult experiments, 8-week-old male mice were administered systemically by tail-vein injection with 1 × 10¹² AAV particles of each monomer [AAV-DdCBE (L) - Nd3-9577 - G1333-C and AAV-DdCBE (H) - Nd3-9577 - G1333-N]. An equal dose was applied in newborn pups (Postnatal day 1—males and females) via the temporal vein, using a 30 G, 30° bevelled needle syringe. Control mice were injected with similar volumes of vehicle buffer (1× PBS, 230 mM NaCl and 5% w/v D-sorbitol).

Genomic DNA isolation and Sanger sequencing of *MT-Nd3* locus. NIH/3T3 mouse cells were collected by trypsinization, washed once in PBS, and resuspended in lysis buffer (1 mM EDTA, 1% Tween 20, 50 mM Tris (pH = 8)) with 200 μg/ml of proteinase K. Lysates were incubated at 56 °C with agitation (300 RPM) for 1 h, and then incubated 95 °C for 10 min before use in downstream applications. Genomic DNA from mouse heart samples (~50 mg) was extracted with a Maxwell[®] 16 Tissue DNA Purification Kit in a Maxwell[®] 16 Instrument (Promega), according to the manufacturer’s instructions.

For Sanger sequencing, the *MT-Nd3* edited region was PCR-amplified with GoTaq G2 DNA polymerase (Promega) using the following primers: Mmu_Nd3_Fw: 5′-GCA TTC TGA CTC CCC CAA AT -3′; and Mmu_Nd3_Rv: 5′-GGC CTA GAG ATA GAA TTG TGA CTA GAA -3′. The PCR was performed with an initial heating step of 1 min at 95 °C followed by 35 cycles of amplification (30 s at 95 °C, 30 s at 63 °C, 30 s at 72 °C), and a final step of 5 min at 72 °C. PCR purification and Sanger sequencing were carried out by Source Bioscience (UK) with the Mmu_Nd3_Rv primer.

High-throughput targeted amplicon sequencing. Genomic DNA was extracted as described above. For high-throughput targeted amplicon resequencing of the *MT-Nd3* region, a 15,781 bp fragment was first amplified by long-range PCR to avoid amplification of nuclear mtDNA pseudogenes (NUMTs), with PrimeSTAR GXL DNA polymerase (TAKARA) using the following primers: Long-R_mtDNA_Fw: 5′-GAG GTG ATG TTT TTG GTA AAC AGG CGG GGT -3′; and LongR_mtDNA_Rv: 5′-GGT TCG TTT GTT CAA CGA TTA AAG TCC TAC GTG -3′. The PCR was performed with an initial heating step of 1 min at 94 °C followed by 10 cycles of amplification (10 s at 98 °C, 13 min at 68 °C), and a final step of 10 min at 72 °C. All PCR reactions from this and the following steps were cleaned up AMPure XP beads (Beckman Coulter, A63881). An aliquot of the purified long-range PCR reactions was amplified with primers containing an overhang adapter sequence, compatible with Illumina index and sequencing primers, in a 25 μl reaction, with Phusion[®] Hot Start Flex 2X Master Mix (NEB, M0536S) using the following primers: Nd3adapter_Fw: 5′-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG TTC TGA ATA AAC CCA GAA GAG AGT -3′; and Nd3adapter_Rv: 5′-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGG TAG ACG TGC AGA GCT TGT -3′. The PCR was performed with an initial heating step of 30 s at 98 °C followed by 12 cycles of amplification (10 s at 98 °C, 15 s at 62.2 °C, 15 s at 72 °C), and a final step of 5 min at 72 °C. The indexing PCR was performed using the Nextera XT Index Kit (Illumina, FC-131-1096) according to the manufacturer’s instructions. The PCR was performed with an initial heating step of 30 s at 98 °C followed by 6 cycles of

amplification (30 s at 98 °C, 30 s at 55 °C, 30 s at 72 °C), and a final step of 5 min at 72 °C. Libraries were subjected to high-throughput sequencing using the Illumina MiSeq platform (PE75) and demultiplexed using the Illumina MiSeq manufacturer's software.

For mtDNA-wide off-targets analysis, two overlapping long amplicons (8331 bp and 8605 bp) covering the full mtDNA molecule were amplified by long-range with PrimeSTAR GXL DNA polymerase (TAKARA) using the following primers: *mmu_ND2_Fw*: 5'-TCT CCG TGC TAC CTA AAC ACC -3'; with *mmu_ND5_Rv*: 5'-GGC TGA GGT GAG GAT AAG CA -3'; and *mmu_ND2_Rv*: 5'-GTA CGA TGG CCA GGA GGA TA -3'; with *mmu_ND5_Fw*: 5'-CTT CCC ACT GTA CAC CAC CA -3'. The PCR was performed with an initial heating step of 1 min at 94 °C followed by 16 cycles of amplification (30 s at 98 °C, 30 s at 60 °C, 9 min at 72 °C), and a final step of 5 min at 72 °C.

For nuclear DNA off-targets assessment, two regions with high homology to the *MT-Nd3* targeted region were analyzed. A region with 100% identity in chromosome 1 was amplified by long-range with PrimeSTAR GXL DNA polymerase (TAKARA) using the following primers: *G40K_NUMT100_Fw2*: 5'-T GC ACT GCT GAC CCA TTA AT -3'; with *G40K_NUMT100_Rv2*: 5'-ACA CAC ACT AGA CAA CAC CCA -3'. The second region with 86% identity in chromosome 14 was amplified using the following primers: *G40K_NUMT86_Fw1*: 5'-CTG GTG GTC ACT TGG TGT GT -3'; with *G40K_NUMT86_Rv1*: 5'-TGT TAC ATG TTT CTC TGT TTT TGC T -3'.

The PCR was performed with an initial heating step of 1 min at 94 °C followed by 20 cycles of amplification (30 s at 98 °C, 30 s at 60 °C, 9 min at 72 °C), and a final step of 5 min at 72 °C. Tagmentation and the indexing PCR were performed using the Nextera XT Index Kit (Illumina, FC-131-1096) according to the manufacturer's instructions and as described above. Libraries were subjected to high-throughput sequencing using the Illumina MiSeq platform (PE250) and demultiplexed using the Illumina MiSeq manufacturer's software.

Processing and mapping of high-throughput data. Quality trimming and 3' end adaptor clipping of sequenced reads were performed simultaneously, using Trim Galore! (-paired)³⁸. For targeted amplicon resequencing of the *MT-Nd3* region and for mtDNA-wide off-targets analysis reads were aligned to ChrM of the mouse reference genome (GRCm38) with Bowtie2 (-very-sensitive; -no-mixed; -no-discordant)³⁹. For nuclear DNA off-targets assessment reads were aligned to the two regions with high homology (GRCm38). Count tables for targeted amplicon resequencing of the *MT-Nd3* region were generated with samtools mpileup (-q 30)⁴⁰ and varscan⁴¹. To study the editing pattern per reading and to determine the percentage of reads that had edits at C₁₂ and C₁₃, we used cutadapt (-e 0; -action=none; -discard-untrimmed)⁴² for the 10nt surrounding the editing site in combination with the samtools flagstat command. For the mtDNA-wide and nuclear DNA off-targets analysis REDtools2.0⁴³ was used (-bq 30) to generate count tables. The average mtDNA-wide C-to-T and G-to-A editing frequency was assessed by summation of all off-target C-to-T and G-to-A editing frequencies divided by the total number of C-to-G sites in mouse mtDNA (5990).

Quantification of viral genomes copy number and relative mtDNA content by quantitative real-time PCR.

All quantitative real-time PCR reactions were performed in a QuantStudio™ 3 system (Thermo Fisher), using a TaqMan™ Gene Expression Master Mix (Thermo Fisher), according to manufacturer's instructions and Ct values obtained using the built-in software. Reactions were performed in a final volume of 20 µl in technical triplicates and each dataset represents samples from an individual mouse. Determination of viral genomes copy number of AAV-DdCBE (L) was performed using primers and probe detecting the DddA₆₀₆ split G1333-C: *VG_1333C_Fw*: 5'-GGA ACT TGC GGA TTT TGT GT -3'; *VG_1333C_Rv*: 5'-TTT GGG AGA ATT GGA GTT GC -3'; *VG_1333C_probe*: 5'-/56-FAM/ CGT AAA ACG /ZEN/ GGG AGC TAC AG /3IABkFQ/ -3'. Copy number of AAV-DdCBE (H) was performed using primers and probe detecting the DddA₆₀₆ split G1333-N: *VG_1333N_Fw*: 5'-CAT ACG CAC TTG GCC CTT AC -3'; *VG_1333N_Rv*: 5'-AAC ACC TTG CTC TCC AGT CC -3'; *VG_1333N_probe*: 5'-/56-FAM/ CAA CTG CCT /ZEN/ GCA TAC AAT GG /3IABkFQ/ -3'. DNA input was normalized to the nuclear gene *Actin*, using the primers and probe: *Mmu_Actin_Fw*: 5'-CTG CTC TTT CCC AGA CGA GG -3'; *Mmu_Actin_Rv*: 5'-AAG GCC ACT TAT CAC CAG CC -3'; *Mmu_Actin_probe*: 5'-/56-TAMN/ ATT GCC TTT CTG ACT AGG TG /3BHQ/ -3'. Standard curves were generated using tenfold serial dilutions of an individual plasmid DNA encoding both G1333-C and G1333-N splits, from 2.5 × 10⁻⁶ to 10⁻³ ng/reaction. Copy numbers were obtained by plotting the logarithm of the plasmid quantity against the measured Ct values from samples.

For relative mtDNA content in heart samples, mtDNA levels were quantified using primers and probe detecting MT-ND1: *Mmu_Nd1_Fw*: 5'-GAG CCT CAA ACT CCA AAT ACT CAC T -3'; *Mmu_Nd1_Rv*: 5'-GAA CTG ATA AAA GGA TAA TAG CTA TGG TTA CTT CA -3'; *Mmu_Nd1_probe*: 5'-/56-FAM/ CCG TAG CCG /ZEN/ AAA CAA T /3IABkFQ/ -3'. Levels of MT-ND1 were normalized to the nuclear gene *Actin*, using the primers and probe: *Mmu_Actin_Fw*: 5'-CTG CTC TTT CCC AGA CGA GG -3'; *Mmu_Actin_Rv*: 5'-AAG GCC ACT TAT CAC CAG CC -3'; *Mmu_Actin_probe*: 5'-/56-TAMN/ ATT GCC TTT CTG ACT AGG TG /3BHQ/ -3'. The Ct values for each sample were used to calculate relative mtDNA copy number using the ΔΔCt analysis, using vehicle-injected mice as control.

Immunoblotting of DdCBEs in mouse hearts. Heart samples (~50 mg) were homogenized in 200 µl of ice-cold RIPA buffer (150 mM NaCl, 1.0% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris, pH 8.0) containing 1X cOmplete™ mini EDTA-free Protease Inhibitor Cocktail (Roche, UK), using a gentleMACS™ dissociator. Homogenates were incubated on ice for 20 mins and then clear by centrifugation (20,000 × g for 20 min at 4 °C). Protein lysates (~20 µg) were mixed with 10 X NuPAGE™ sample reducing agent and 4X NuPAGE™ LDS sample buffer (Invitrogen), and incubated for 5 min at 95 °C. The boiled protein samples were then separated in a Bolt 4–12% Bis-Tris (Thermo Fisher) pre-cast gel and later transferred to a PVDF membrane using an iBlot 2 gel transfer system (Thermo Fisher), according to the manufacturer's recommendations. The residual proteins that remained in the gel were detected using SimpleBlue SafeStain (Thermo Fisher) and used as a loading control. The membrane was blocked in 5% milk in PBS with 0.1% Tween 20 (PBS-T) for 1 h at room temperature (RT) and then incubated with either rat anti-HA-tag antibody (Roche, 11867423001), 1:1000 diluted in 5% milk in PBST or mouse anti-FLAG-M2-tag (Sigma-Aldrich, F3165), 1:2000 diluted in 5% milk in PBST. Membranes were washed three times with PBS-T for 10 min at RT and then incubated with HRP-linked secondary antibodies, either anti-rat IgG (Cell Signaling, 7077 S) or anti-mouse IgG (Promega, W4021), 1:5000 diluted in 5% milk in PBST. The membranes were washed another three times as above and imaged digitally with an Amersham Imager 680 blot and gel imager (GE Healthcare), upon incubation with Amersham ECL™ Western Blotting Detection Reagents (GE Healthcare).

Immunohistochemistry and microscopy. At sacrifice, mouse heart samples were frozen by immersion in isopentane cooled in liquid nitrogen. For immunohistochemistry analysis, heart samples were mounted in optimal cutting temperature compound (OCT) and sectioned on a cryostat at -20 °C to a thickness of 8 µm. Sections were fixed in 10% neutral buffered formalin (SIGMA, HT501128) at RT. After three washes in PBS, samples were permeabilized using 0.2% Triton X-100 in PBS for 15 min at RT, followed by three washes in PBS, and then blocked in 5% normal goat serum + 2% BSA in PBS for 1 h at RT. After three washes in PBS, samples were incubated with the rabbit anti-HA-Tag (Cell Signaling, 3724) primary antibody 1:2000 in DAKO solution (Agilent, S3022) for 1 h at RT, followed by three additional washes in PBS. Sections were then incubated with goat anti-rabbit IgG (H + L) AlexaFluor diluted 1:300 in DAKO diluent solution 568 for 1 h at RT, washes three times in PBS and finally coverslip with prolonged diamond antifade mountant with DAPI (Thermo Fisher, P36962). All confocal images were acquired using a Zeiss LSM880 microscope using identical acquisition parameters.

Statistics. Graphical visualization of data and all statistical analyses were performed with GraphPad Prism software (version 8.0). All numerical data are expressed as mean ± standard error of mean (SEM). Ordinary one-way ANOVA with Dunnett's test was used for multiple comparisons. Animals were randomized and no blinding to the operator was used.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files, apart from proprietary data which are available upon request by contacting the authors. The NGS files generated in this study have been deposited in GEO: GSE184064. Source data are provided with this paper.

Code availability

On and off-target effects from next-generation sequencing data were calculated with Varscan2, source code: <https://github.com/jedje/varsan2>, written by Koboldt et al.⁴¹.

Received: 14 September 2021; Accepted: 21 January 2022;

Published online: 08 February 2022

References

- Vafai, S. B. & Mootha, V. K. Mitochondrial disorders as windows into an ancient organelle. *Nature* **491**, 374–383 (2012).
- Suomalainen, A. & Battersby, B. J. Mitochondrial diseases: the contribution of organelle stress responses to pathology. *Nat. Rev. Mol. Cell Biol.* **19**, 77–92 (2018).
- Rath, S. et al. MitoCarta3.0: an updated mitochondrial proteome now with sub-organelle localization and pathway annotations. *Nucleic Acids Res.* **49**, D1541–D1547 (2021).
- Gorman, G. S. et al. Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease. *Ann. Neurol.* **77**, 753–759 (2015).
- Russell, O. M., Gorman, G. S., Lightowlers, R. N. & Turnbull, D. M. Mitochondrial diseases: hope for the future. *Cell* **181**, 168–188 (2020).

6. Kauppi, T. E. S., Kauppi, J. H. K. & Larsson, N. G. Mammalian mitochondria and aging: an update. *Cell Metab.* **25**, 57–71 (2017).
7. Yonova-Doing, E. et al. An atlas of mitochondrial DNA genotype-phenotype associations in the UK Biobank. *Nat. Genet.* **53**, 982–993 (2021).
8. Filigrana, R., Menzani, M., Alsina, D. & Larsson, N. G. Mitochondrial DNA copy number in human disease: the more the better? *FERS Lett.* **595**, 976–1002 (2021).
9. Stewart, J. B. & Chinnery, P. F. Extreme heterogeneity of human mitochondrial DNA from organelles to populations. *Nat. Rev. Genet.* **22**, 106–118 (2021).
10. Gorman, G. S. et al. Mitochondrial diseases. *Nat. Rev. Dis. Prim.* **2**, 16080 (2016).
11. Gammage, P. A., Moraes, C. T. & Minczuk, M. Mitochondrial genome engineering: the revolution may not be CRISPR-led. *Trends Genet.* **34**, 101–110 (2018).
12. Srivastava, S. & Moraes, C. T. Manipulating mitochondrial DNA heteroplasmy by a mitochondrially targeted restriction endonuclease. *Hum. Mol. Genet.* **10**, 3093–3099 (2001).
13. Tanaka, M. et al. Gene therapy for mitochondrial disease by delivering restriction endonuclease SmaI into mitochondria. *J. Biomed. Sci.* **9**, 534–541 (2002).
14. Alexeyev, M. F. et al. Selective elimination of mutant mitochondrial genomes as therapeutic strategy for the treatment of NARP and MILS syndromes. *Gene Ther.* **15**, 516–523 (2008).
15. Minczuk, M., Papworth, M. A., Miller, J. C., Murphy, M. P. & Klug, A. Development of a single-chain, quasi-dimeric zinc-finger nuclease for the selective degradation of mutated human mitochondrial DNA. *Nucleic Acids Res.* **36**, 3926–3938 (2008).
16. Minczuk, M., Kolasinska-Zotwierz, P., Murphy, M. P. & Papworth, M. A. Construction and testing of engineered zinc-finger proteins for sequence-specific modification of mtDNA. *Nat. Protoc.* **5**, 342–356 (2010).
17. Bacman, S. R., Williams, S. L., Pinto, M., Peralta, S. & Moraes, C. T. Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs. *Nat. Med.* **19**, 1111–1113 (2013).
18. Gammage, P. A., Rorbach, J., Vincent, A. I., Rebar, E. J. & Minczuk, M. Mitochondrially targeted ZFNs for selective degradation of pathogenic mitochondrial genomes bearing large-scale deletions or point mutations. *EMBO Mol. Med.* **6**, 458–466 (2014).
19. Hashimoto, M. et al. MitoTALEN: A general approach to reduce mutant mtDNA loads and restore oxidative phosphorylation function in mitochondrial diseases. *Mol. Ther.* **23**, 1592–1599 (2015).
20. Gammage, P. A. et al. Near-complete elimination of mutant mtDNA by iterative or dynamic dose-controlled treatment with mtZFNs. *Nucleic Acids Res.* **44**, 7804–7816 (2016).
21. Yang, Y. et al. Targeted elimination of mutant mitochondrial DNA in MELAS-iPSCs by mitoTALENs. *Protein Cell* **9**, 283–297 (2018).
22. Silva-Pinheiro, P. & Minczuk, M. The potential of mitochondrial genome engineering. *Nat. Rev. Genet.* <https://doi.org/10.1038/s41576-021-00432-x> (2021).
23. Bayona-Bafalay, M. P., Blits, B., Battersby, B. J., Shoubridge, E. A. & Moraes, C. T. Rapid directional shift of mitochondrial DNA heteroplasmy in animal tissues by a mitochondrially targeted restriction endonuclease. *Proc. Natl Acad. Sci. USA* **102**, 14392–14397 (2005).
24. Bacman, S. R., Williams, S. L., Duan, D. & Moraes, C. T. Manipulation of mtDNA heteroplasmy in all striated muscles of newborn mice by AAV9-mediated delivery of a mitochondria-targeted restriction endonuclease. *Gene Ther.* **19**, 1101–1106 (2012).
25. Bacman, S. R. et al. MitoTALEN reduces mutant mtDNA load and restores tRNA(Ala) levels in a mouse model of heteroplasmic mtDNA mutation. *Nat. Med.* **24**, 1696–1700 (2018).
26. Gammage, P. A. et al. Genome editing in mitochondria corrects a pathogenic mtDNA mutation in vivo. *Nat. Med.* **24**, 1691–1695 (2018).
27. Zekonyte, U. et al. Mitochondrial targeted meganuclease as a platform to eliminate mutant mtDNA in vivo. *Nat. Commun.* **12**, 3210 (2021).
28. Mok, B. Y. et al. A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. *Nature* **583**, 631–637 (2020).
29. Lee, H. et al. Mitochondrial DNA editing in mice with DddA-TALE fusion deaminases. *Nat. Commun.* **12**, 1190 (2021).
30. Guo, J. et al. Precision modeling of mitochondrial diseases in zebrafish via DdCBE-mediated mtDNA base editing. *Cell Discov.* **7**, 78 (2021).
31. Agip, A. A. et al. Cryo-EM structures of complex I from mouse heart mitochondria in two biochemically defined states. *Nat. Struct. Mol. Biol.* **25**, 548–556 (2018).
32. Blaza, J. N., Vinothkumar, K. R. & Hirst, J. Structure of the inactive state of mammalian respiratory complex I. *Structure* **26**, 312–319 e313 (2018).
33. Gross, N. J., Getz, G. S. & Rabinowitz, M. Apparent turnover of mitochondrial deoxyribonucleic acid and mitochondrial phospholipids in the tissues of the rat. *J. Biol. Chem.* **244**, 1552–1562 (1969).
34. Silva Ramos, E. et al. Mitochondrial fusion is required for regulation of mitochondrial DNA replication. *PLoS Genet.* **15**, e1008085 (2019).
35. Stewart, J. B. Current progress with mammalian models of mitochondrial DNA disease. *J. Inher. Metab. Dis.* **44**, 325–342 (2021).
36. Kauppi, J. H. K. et al. A phenotype-driven approach to generate mouse models with pathogenic mtDNA mutations causing mitochondrial disease. *Cell Rep.* **16**, 2980–2990 (2016).
37. Chouchani, E. T. et al. Cardioprotection by S-nitrosylation of a cysteine switch on mitochondrial complex I. *Nat. Med.* **19**, 753–759 (2013).
38. Krueger, F., James, F., Ewels, P., Aiyounian, E. & Schuster-Boeckler, B. FelixKrueger/TrimsGalore: v0.6.7. *Zenodo* <https://doi.org/10.5281/zenodo.5127899> (2021).
39. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
40. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* <https://doi.org/10.1093/gigascience/gjab008> (2021).
41. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
42. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBio* **17**, 10–12 (2011).
43. Flati, T. et al. HPC-REDTools: a novel HPC-aware tool for improved large-scale RNA-editing analysis. *BMC Bioinformatics* **21**, 353 (2020).

Acknowledgements

We would like to acknowledge the members of the Mitochondrial Genetics Group (MRC-MBU, University of Cambridge) for useful discussion during the course of this research. We would like to thank Ahmed-Noor Agip for his help with Fig. 1c. This work was supported by core funding from Medical Research Council UK (MC_UU_00015/4). P.S.-P. and P.A.N. are additionally supported by The Champ Foundation and The Lily Foundation, respectively.

Author contributions

P.S.-P. and M.M. planned and designed experiments; P.A.N. and C.D.M. performed the immunohistochemistry and NGS experiments; L.V.H. and P.S.-P. analyzed the NGS experiments; K.T. managed the animal work and performed the histology experiments; P.S.-P. performed the remaining experiments. P.S.-P. and M.M. drafted the manuscript. M.M. supervised the study. All authors revised the manuscript.

Competing interests

M.M. is a co-founder, shareholder and member of the Scientific Advisory Board of Pretzel Therapeutics, Inc. P.S.-P., P.A.N., and C.D.M. provide consultancy services for Pretzel Therapeutics, Inc. The remaining authors declare no competing interests.

Additional information


Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-28358-w>.

Correspondence and requests for materials should be addressed to Michal Minczuk.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022